THE SCALING MEAN AND A LAW OF LARGE PERMANENTS

JAIRO BOCHI, GODOFREDO IOMMI, AND MARIO PONCE

ABSTRACT. In this paper we study two types of means of the entries of a non-negative matrix: the *permanental mean*, which is defined using permanents, and the *scaling mean*, which is defined in terms of an optimization problem. We explore relations between these two means, making use of important results by Ergorychev and Falikman (the van der Waerden conjecture), Friedland, Sinkhorn, and others. We also define a scaling mean for functions in a much more general context. Our main result is a Law of Large Permanents, a point-wise ergodic theorem for permanental means of dynamically defined matrices that expresses the limit as a functional scaling mean. The concepts introduced in this paper are general enough so to include as particular cases certain classical types of means, as for example symmetric means and Muirhead means. As a corollary, we reobtain a formula of Halász and Székely for the limit of the symmetric means of a stationary random process.

1. INTRODUCTION

1.1. Matrix means. Among the myriad notions of *means* that appear in mathematics, the arithmetic and the geometric means are the most important ones. In this paper we introduce two notions of mean of the entries of a matrix. Both notions are obtained by combination of the arithmetic and the geometric means. Let us define them now, leaving details and proofs for later.

The first notion is the *permanental mean* of an $n \times n$ matrix A with nonnegative entries, which is defined as the n-th root of the arithmetic mean of the products of the diagonals of A. The reason for the name is that this mean is related to the permanent of A:

$$\mathrm{pm}(A) = \left(\frac{\mathrm{per}\, A}{n!} \right)^{1/n} .$$

We discuss the main properties of the permanental mean in § 2.1. Some classic types of means are particular cases of the permanental mean: see Section 5.

Two matrices A and B of the same size are called *scalings* of one another if their entries are related by $b_{ij} = x_i a_{ij} y_j$ for suitable vectors x, y with positive entries x_i, y_j. If the geometric means of the entries of both vectors are 1, we say that B is a *normalized scaling* of A. Assuming A to have nonnegative entries, we define the *scaling mean* of A, denoted $\mathrm{sm}(A)$, as the infimum of the arithmetic means of all normalized scalings of A. It turns out that the scaling mean shares many properties with the permanental mean, and it can be characterized in several other ways: see §§ 2.2–2.3.

This paper is devoted to the study of the relations between the scaling mean and the permanental mean in a range of different settings.

If A is a doubly stochastic $n \times n$ matrix, i.e. the arithmetic mean of each row and each column is $1/n$, then by the celebrated van der Waerden conjecture (which was an open problem for more than 50 years until being proved in 1981), the permanent of A is at least $n!/n^n$ or, in other words, its permanental mean is at least $1/n$. On the other hand, as a consequence of the AM-GM inequality (see § 2.2), the scaling mean of A equals $1/n$, and therefore

$$\mathrm{pm}(A) \geqslant \mathrm{sm}(A) \tag{1.1}$$

for doubly stochastic matrices. It transpires that this inequality holds true for *all* nonnegative square matrices. This generalization of the van der Warden conjecture is actually a consequence of it, together with the fact that "most" nonnegative matrices (in particular all positive ones) are scalings of doubly stochastic matrices, a discovery that dates back from the 1960's with the works of Sinkhorn and others. Details are provided in §§ 2.3–2.5, where we also characterize the cases of equality in (1.1). Moreover, and more importantly for the purposes of this paper, approximate equality holds in (1.1) for certain "repetitive" large matrices: The precise statement is given in § 2.6 and is basically a reinterpretation of a result by Friedland, who in 1979 proved an asymptotic weaker version of the van der Waerden conjecture. That fact is one of the foundations for the main result of this paper, which we describe next.

1.2. **The general setting for scaling problems.** Scaling problems have been studied in infinite dimensions, for infinite matrices and for functions. In this paper we introduce a more general and abstract setting that includes the above ones as particular cases and opens a door for other possibilities. Basically, we consider measurable functions on arbitrary probability spaces, and the allowed scaling functions are those that are measurable with respect to certain fixed sub-σ-algebras; see Section 3 for details.

We extend the definition of scaling mean to this general setting. We also prove the existence of "doubly stochastic" scalings under certain boundedness assumptions, thus extending previously known facts about the so-called Sinkhorn decompositions for matrices or DAD problems for functions.

1.3. **A Law of Large Permanents.** On the other hand, it is not clear what should be the permanental mean of an *infinite* matrix. A natural attempt is to consider square truncations and then take the limit of the corresponding permanental means as the size of the square tends to infinity, provided of course that such limit exists. We prove that this is indeed the case for some classes of dynamically defined matrices. Moreover, we identify the limit permanental mean as the scaling mean of the function that controls the distribution of the matrix entries. The precise result is as follows:

Theorem 4.1. *Let (X, μ), (Y, ν) be Lebesgue probability spaces, let $T \colon X \to X$ and $S \colon Y \to Y$ be ergodic measure preserving transformations, and let $f \colon X \times Y \to \mathbb{R}$ be a positive measurable function essentially bounded away from zero and infinity.*

Then for $\mu \times \nu$-almost every $(x, y) \in X \times Y$, the permanental mean of the matrix

$$\begin{pmatrix} f(x,y) & f(Tx,y) & \cdots & f(T^{n-1}x,y) \\ f(x,Sy) & f(Tx,Sy) & \cdots & f(T^{n-1}x,Sy) \\ \vdots & \vdots & & \vdots \\ f(x,S^{n-1}y) & f(Tx,S^{n-1}y) & \cdots & f(T^{n-1}x,S^{n-1}y) \end{pmatrix}$$

converges as $n \to \infty$ to the scaling mean of the function f.

In this fairly general pointwise ergodic theorem we not only prove the almost everywhere convergence but we also identify the limit, thus completely solving the two main questions that arise when considering ergodic averages.

Several results scattered in the literature are contained in this *Law of Large Permanents*. For example, the aforementioned result of Friedland is a corollary. The Law is flexible enough so that we can deduce from it variants of Birkhoff's ergodic theorem where the arithmetic means are replaced by other types of means. For example, in § 5.1 we deduce an ergodic theorem for symmetric means, thus reobtaining in a more transparent way a formula due to Halász and Székely [HS].

Let us remark that the literature contains asymptotic results about permanents of random oblong (i.e. non-square) matrices, a subject we will not deal with: see [RW] and references therein. There are fewer results for square matrices: we can only cite [TV].

1.4. **Organization of the paper.** In Section 2 we study permanental and scaling means of matrices, proving the properties mentioned above, among several others. In Section 3 we introduce an abstract setting for scaling problems, where we define scaling means for functions and also prove an existence theorem for functional scaling. Section 4 is devoted to the proof of the Law of Large Permanents (Theorem 4.1). In Section 5 we present some corollaries concerning symmetric and Muirhead means. Section 6 discusses some of the questions arising from our results, and poses some conjectures.

Throughout this paper we use the following notations:

$$\mathbb{R}_+ := [0, \infty), \quad \mathbb{R}_{++} := (0, \infty).$$

The set of real (resp., nonnegative, positive) $m \times n$ matrices is denoted by $\mathbb{R}^{m \times n}$ (resp., $\mathbb{R}_+^{m \times n}$, $\mathbb{R}_{++}^{m \times n}$).

2. Permanental and scaling means of matrices

2.1. **The permanental mean.** The *permanent* of a square matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ is defined as

$$\operatorname{per} A := \sum_\sigma \prod_{i=1}^n a_{i,\sigma(i)},$$

where σ runs on the permutations of $\{1, \ldots, n\}$. This function was introduced independently by Cauchy and Binet around 1812. It has many applications in combinatorics [vLW], probability [Ba3], among other areas. The book [Mi] is wholly dedicated to permanents and contains historical information and the most relevant results and applications up to the late 1970's.

As the determinant, the permanent is a symmetric multilinear function of the rows (or columns) of a matrix; it is also invariant under transposition. Despite the similarities between the definitions of determinant and permanent, there is no

permanental analog of the Gaussian elimination algorithm, and indeed the evaluation of the permanent is a computationally much more complex problem: see [BR, pp. 245ff].

When dealing with nonnegative matrices, the following function is in some senses better behaved than the permanent itself:

Definition 2.1. The *permanental mean* of a square nonnegative matrix $A \in \mathbb{R}_+^{n \times n}$ is defined as

$$\mathrm{pm}(A) := \left(\frac{\mathrm{per}\, A}{n!} \right)^{1/n} .$$

The permanental mean has the following properties, which allow us to consider it as a sort of average of the entries of the matrix:

- *Monotonicity*: the permanental mean is increasing as a function of each of the entries of the matrix.
- *Reflexivity*: a matrix with constant entries has a permanental mean equal to this constant.

These two properties imply the following one:

- *Internality*: the permanental mean is between the minimum and the maximum of the entries of the matrix.

Some additional properties are:

- *Continuity*.
- *Row-wise and column-wise homogeneity*: the permanental mean on $\mathbb{R}_+^{n \times n}$ is homogeneous of degree $1/n$ as a function on each row and each column, and in particular it is a homogenous function of degree 1. Equivalently, if A is a nonnegative square matrix and D is a diagonal matrix of the same size and with positive main diagonal, then

$$\mathrm{pm}(AD) = \mathrm{pm}(DA) = \mathrm{gm}(D)\,\mathrm{pm}(A)\,,$$

  where $\mathrm{gm}(D)$ denotes the geometric mean of the entries along the main diagonal of D.
- *Row-wise, column-wise, and transpositional symmetry*: the permanental mean is invariant under permutations of rows or columns, and under transposition of the matrix. Equivalently, if A is a nonnegative square matrix and P is a permutation matrix of the same size then

$$\mathrm{pm}(AP) = \mathrm{pm}(PA) = \mathrm{pm}(A^{\top}) = \mathrm{pm}(A)\,.$$

A square matrix with nonnegative entries is called *doubly stochastic* if the sums of the entries on each row and each column are equal to 1. The set of $n \times n$ doubly stochastic matrices is denoted by Ω_n. By a classical theorem of Garrett Birkhoff, Ω_n is a convex polytope whose vertices are the permutation matrices: see e.g. [BR, MM, Mi]. A great deal of work has been devoted to study the permanent of this kind of matrices, the van der Waerden conjecture being probably the most relevant topic. If $S \in \Omega_n$ then

$$\frac{n!}{n^n} \leqslant \mathrm{per}\, S \leqslant 1\,. \tag{2.1}$$

The upper bound is trivial, while the lower bound was conjectured by van der Waerden [vdW] in 1926 and proved in 1981 independently by Egorychev [Eg] and Falikman [Fa]. Moreover, the minimum of the permanent on Ω_n is attained at the

matrix J_n all of whose entries are equal to $1/n$ and the maximum is attained at permutation matrices. In terms of the permanental mean, theses bounds become:

$$\frac{1}{n} \leqslant \mathrm{pm}(S) \leqslant \frac{1}{(n!)^{1/n}} \quad \text{if } S \in \Omega_n. \tag{2.2}$$

2.2. **The scaling mean.** In the following definition, we identify each $x \in \mathbb{R}^n$ with a column vector, and denote by $\mathrm{gm}(x)$ denotes the geometric mean of its entries.

Definition 2.2. The *scaling mean* of a nonnegative matrix $A \in \mathbb{R}_+^{m \times n}$ is

$$\mathrm{sm}(A) := \frac{1}{mn} \inf_{x,y} \frac{x^\top A y}{\mathrm{gm}(x)\,\mathrm{gm}(y)}\,, \tag{2.3}$$

where x runs on $\mathbb{R}_{++}^n$ and y runs on $\mathbb{R}_{++}^m$.

Though the definition above also applies to oblong matrices, in the sequel we will mostly consider square matrices.

As the permanental mean, the scaling mean has the properties of monotonicity, reflexivity (a consequence of the AM-GM inequality), and internality, which is why we call it a "mean".

The scaling mean also has the properties of row-/column-wise homogeneity and row-/column-wise symmetry; these follow from the analogous properties of the geometric mean. Transpositional symmetry obviously holds. Other useful properties are given by the following three propositions:

Proposition 2.3. *The function* $\mathrm{sm}\colon \mathbb{R}_+^{m \times n} \to \mathbb{R}_+$ *is concave and continuous.*

Proof. Being defined as the infimum of a family of linear functions, the scaling mean is a concave and upper semicontinuous function. Let us prove lower semicontinuity at an arbitrary fixed $A = (a_{ij}) \in \mathbb{R}_+^{m \times n}$. Let H be the set of nonnegative matrices that have the same pattern of zeros as A, i.e.,

$$H := \left\{ (b_{ij}) \in \mathbb{R}_+^{m \times n} \,:\, b_{ij} = 0 \iff a_{ij} = 0 \right\}.$$

Since H is convex and relatively open with respect to its affine hull in $\mathbb{R}^{m \times n}$, it follows from concavity (see [Ro, Thrm. 10.1]) that the restriction of the function sm to H is continuous. Consider a sequence $\left(A_k = (a_{k,ij}) \right)$ in $\mathbb{R}_+^{m \times n}$ converging to A. Define $B_k = (a_{k,ij})$ by

$$b_{k,ij} := \begin{cases} a_{k,ij} & \text{if } a_{ij} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

By monotonicity, $\mathrm{sm}(B_k) \leqslant \mathrm{sm}(A_k)$. Moreover, $B_k \to A$ and $B_k \in H$ for large enough k. Therefore $\liminf \mathrm{sm}(A_k) \geqslant \lim \mathrm{sm}(B_k) = \mathrm{sm}(A)$, thus establishing lower semicontinuity. $\qquad\square$

Recall that $\Omega_n \subset \mathbb{R}_+^{n \times n}$ is the set of doubly stochastic matrices.

Proposition 2.4. *If* $A \in \Omega_n$ *then* $\mathrm{sm}(A) = 1/n$.

Proof. Let $A \in \Omega_n$. For all $x, y \in \mathbb{R}_{++}^n$, by the weighted AM-GM inequality,

$$\frac{1}{n} x^\top A y = \sum_{i,j} \frac{a_{ij}}{n}\, x_i y_j \geqslant \prod_{i,j} (x_i y_j)^{a_{ij}/n} = \mathrm{gm}(x)\,\mathrm{gm}(y).$$

Equality is attained if $x = y = (1, \ldots, 1)$, say. Therefore $\mathrm{sm}(A) = 1/n$. $\qquad\square$

Proposition 2.5. *If A is nonnegative square matrix of block triangular form $A = \left(\begin{smallmatrix} B & R \\ 0 & C \end{smallmatrix} \right)$ where B and C are square matrices, then $\mathrm{sm}(A)$ does not depend on R.*

Proof. Suppose A, B, and C have sizes $n \times n$, $k \times k$, and $\ell \times \ell$ respectively, so $n = k + \ell$. For all $u \in \mathbb{R}_{++}$, consider the matrix:

$$D(u) := \mathrm{diag}\,\big(\underbrace{u^{1/k}, \ldots, u^{1/k}}_{k}, \underbrace{u^{-1/\ell}, \ldots, u^{-1/\ell}}_{\ell}\big).$$

Given $x, y \in \mathbb{R}^n_{++}$, let $x_u := D(u)x$ and $y_u := D(u^{-1})y$. Then

$$\mathrm{gm}(x_u) = \mathrm{gm}(x), \quad \mathrm{gm}(y_u) = \mathrm{gm}(y), \quad \text{and} \quad x_u^\top A y_u = x^\top \begin{pmatrix} B & u^{1/k+1/\ell}R \\ 0 & C \end{pmatrix} y.$$

Therefore

$$\mathrm{sm}(A) \leqslant \frac{1}{n^2} \inf_u \frac{x_u^\top A y_u}{\mathrm{gm}(x_u)\,\mathrm{gm}(y_u)} = \frac{1}{n^2} \frac{x^\top A_0 y}{\mathrm{gm}(x)\,\mathrm{gm}(y)}\,,$$

where A_0 is the matrix obtained from A by replacing R with 0. Taking infimum on x, y, we obtain $\mathrm{sm}(A) \leqslant \mathrm{sm}(A_0)$. On the other hand, $\mathrm{sm}(A) \geqslant \mathrm{sm}(A_0)$ by monotonicity, thus concluding the proof. $\qquad\square$

Remark 2.6. As a consequence of the AM-GM inequality, we have:

$$z \in \mathbb{R}^n_+ \;\Rightarrow\; \inf_{x \in \mathbb{R}^n_{++}} \frac{x^\top z}{\mathrm{gm}(x)} = n\,\mathrm{gm}(z).$$

Therefore (2.3) can be rewritten as:

$$A \in \mathbb{R}^{m \times n}_+ \;\Rightarrow\; \mathrm{sm}(A) = \frac{1}{m} \inf_{y \in \mathbb{R}^m_{++}} \frac{\mathrm{gm}(Ay)}{\mathrm{gm}(y)}\,, \tag{2.4}$$

As immediate consequences of this formula, we obtain the following additional properties of the scaling mean:

$$A \in \mathbb{R}^{m \times n}_+,\ B \in \mathbb{R}^{n \times r}_+ \;\Rightarrow\; \mathrm{sm}(AB) \geqslant n\,\mathrm{sm}(A)\,\mathrm{sm}(B), \tag{2.5}$$

$$A \in \mathbb{R}^{n \times n}_+ \;\Rightarrow\; \mathrm{sm}(A) \leqslant \frac{\rho(A)}{n}\,, \tag{2.6}$$

where ρ denotes spectral radius. See also Remarks 2.12 and 2.21.

Remark 2.7. The expressions after the inf's in (2.3) and (2.4) are log-log-convex functions with respect to (x, y) and y, respectively, so the computation of the scaling mean from either formula is equivalent to a convex minimization problem.

Remark 2.8. One may ask if the permanental mean also has the concavity property obtained for the scaling mean in Proposition 2.3. The answer is no, because otherwise the minimum of permanental mean on the convex polytope Ω_n would be attained at the boundary, while we know by the Egorychev–Falikman theorem that this is not the case. On the other hand, the permanental mean is indeed concave among positive definite symmetric matrices: see [Bh, p. 282].

2.3. **Matrix scaling and Sinkhorn decompositions.** Recall from the introduction that two matrices A, $B \in \mathbb{R}^{m \times n}$ are called *scalings* of one another if there exist diagonal matrices $E \in \mathbb{R}^{m \times m}$, $D \in \mathbb{R}^{n \times n}$ with positive main diagonals such that $A = EBD$. Matrix scaling has applications in numerical analysis [GvL] and economics [LS]. We will be especially interested in scaling when one of the matrices is doubly stochastic, so we introduce the following:

Definition 2.9. A *Sinkhorn decomposition* of a square matrix A is a factorization of the form

$$A = DSE \quad \text{where} \begin{cases} S \text{ is doubly stochastic,} \\ D, E \text{ are diagonal with positive main diagonals.} \end{cases}$$

In 1964, Sinkhorn [Si1] proved that any positive square matrix can be scaled to a doubly stochastic matrix and moreover the corresponding Sinkhorn decomposition is unique up to multiplication of D and E by positive factors λ and λ^{-1}, respectively. Not all nonnegative square matrices possess Sinkhorn decompositions, however; an example is $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

The following proposition relates the previously defined means and Sinkhorn decompositions. Recall that if D is a diagonal matrix with positive entries along the main diagonal then $\operatorname{gm}(D)$ denotes the geometric mean of these entries.

Proposition 2.10. *If $A \in \mathbb{R}_+^{n \times n}$ has a Sinkhorn decomposition $A = DSE$ then*

$$\operatorname{sm}(A) = \frac{1}{n} \operatorname{gm}(D) \operatorname{gm}(E) = \frac{1}{n} \frac{\operatorname{pm}(A)}{\operatorname{pm}(S)}. \tag{2.7}$$

Moreover, if x and y are the vectors whose entries form the diagonals of D^{-1} and E^{-1}, respectively, then the infimum in (2.3) is attained at x, y.

Proof. The first part is an immediate consequence of Proposition 2.4 and the row-/column-wise homogeneity of the scaling and the permanental means. If x and y are as described, then

$$\frac{1}{n^2} \frac{x^\top A y}{\operatorname{gm}(x) \operatorname{gm}(y)} = \frac{1}{n} \operatorname{gm}(D) \operatorname{gm}(E),$$

so the second part follows from the first one. $\qquad \square$

Example 2.11. The Sinkhorn decomposition of a 2×2 positive matrix is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{ad}+\sqrt{bc}}{\sqrt{cd}} & 0 \\ 0 & \frac{\sqrt{ad}+\sqrt{bc}}{\sqrt{ab}} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{ad}}{\sqrt{ad}+\sqrt{bc}} & \frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}} \\ \frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}} & \frac{\sqrt{ad}}{\sqrt{ad}+\sqrt{bc}} \end{pmatrix} \begin{pmatrix} \sqrt{ac} & 0 \\ 0 & \sqrt{bd} \end{pmatrix},$$

and therefore the scaling mean is

$$\operatorname{sm} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{\sqrt{ad} + \sqrt{bc}}{2}.$$

By continuity (Proposition 2.3), this formula also holds for nonnegative matrices.

Remark 2.12. Relations (2.4) and (2.6), once rewritten using the first equality from (2.7), were stated explicitly and proved by London [Lo].

2.4. **Existence of Sinkhorn decompositions.** We need some definitions:

- Two matrices of the same size have the *same zero pattern* if their zero entries occupy the same positions.
- A *diagonal* of a square matrix is a sequence of entries containing exactly one entry from each row and one entry from each column; a diagonal is called *positive* if each of its elements is positive.
- Two matrices A, $B \in \mathbb{R}^{m \times n}$ are *permutationally equivalent* if there exist permutation matrices $P \in \mathbb{R}^{m \times m}$, $Q \in \mathbb{R}^{n \times n}$ such that $B = PAQ$.
- A matrix $A \in \mathbb{R}^{n \times n}$ called *fully indecomposable* if either $n = 1$ and $A \neq 0$ or $n \geqslant 2$ and A is not permutationally equivalent to a matrix of the form $\left(\begin{smallmatrix} B & R \\ 0 & C \end{smallmatrix} \right)$, where B and C are square matrices.
- The *direct sum* of two square matrices A, B is the matrix $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix} \right)$; this is an associative (but non commutative) operation.

Theorem 2.13 (Perfect–Mirsky [PM]). *Let $A \in \mathbb{R}_+^{n \times n}$. The following assertions are equivalent:*

(a) *A has the zero pattern of some doubly stochastic matrix;*
(b) *every positive entry of A lies on a positive diagonal;*
(c) *A is permutationally equivalent to a direct sum of fully indecomposable matrices.*

Moreover, if $n \geqslant 2$ then the assertions above are equivalent to:

(d) *A is not permutationally equivalent to a matrix of the form $\left(\begin{smallmatrix} B & R \\ 0 & C \end{smallmatrix} \right)$, where the matrices B and C are square and R is nonzero.*

Definition 2.14. Let $\mathcal{P}_n$ be the subset of $\mathbb{R}_+^{n \times n}$ formed by matrices that satisfy any of the equivalent conditions in Theorem 2.13.

The following theorem provides still another equivalent definition for the set $\mathcal{P}_n$:

Theorem 2.15 (General Sinkhorn decompositions). *A matrix $A \in \mathbb{R}_+^{n \times n}$ has a Sinkhorn decomposition $A = DSE$ if and only if $A \in \mathcal{P}_n$, and in that case the doubly stochastic matrix S is unique. Moreover, the map $A \mapsto S$ is continuous.*

The first part of the theorem was proved independently and by means of different arguments by several authors [BPS, SK, MO]. Continuity was proved by Sinkhorn [Si2] (see also [Me2]). Since these early papers, many other proofs, generalizations, and numerical studies have appeared in the literature.

Some of the proofs of existence of Sinkhorn decompositions are closely related to the scaling mean (and indeed motivated our Definition 2.2), so let us sketch them.

In order to prove that matrices in $\mathcal{P}_n$ have Sinkhorn decompositions, it is sufficient to consider fully indecomposable matrices A, because then the general case follows by characterization (c) in Theorem 2.13. In that case, Marshall and Olkin [MO] have shown that the function after the inf in (2.3) diverges to infinity as (x, y) approaches the boundary of the cone $\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$. In particular, the infimum is attained at some (x, y) in the interior of the cone. Then a Lagrange multipliers calculation shows that:

$$\left. \begin{array}{l} \lambda := x^\top A y / n \\ D := \lambda \operatorname{diag}(x_1^{-1}, \ldots, x_n^{-1}) \\ E := \operatorname{diag}(y_1^{-1}, \ldots, y_n^{-1}) \end{array} \right\} \Rightarrow S := D^{-1} A E^{-1} \text{ is doubly stochastic,}$$

and so $A = DSE$ is the sought after Sinkhorn decomposition. Notice that this strategy is basically a converse of Proposition 2.10. Djoković [Dj] and London [Lo] independently discovered an analogous proof based instead on the minimization problem (2.4). For a related discussion, see [Ba2, § 2].

Let us introduce a technical device that will be useful later. Given $A \in \mathbb{R}_+^{n \times n}$, let $\Pi(A)$ denote the matrix obtained by keeping all entries that lie on positive diagonals, and setting the remaining entries to zero. The map Π is a projection, i.e. $\Pi \circ \Pi = \Pi$, and although it is discontinuous it has useful properties:

Proposition 2.16. *If $A \in \mathbb{R}_+^{n \times n}$ then:*

    (a) $\Pi(A) \in \mathcal{P}_n \cup \{0\}$;
    (b) $\Pi(A) = 0$ *iff* $\operatorname{per} A = 0$;
    (c) $\operatorname{pm} \Pi(A) = \operatorname{pm}(A)$;
    (d) $\operatorname{sm} \Pi(A) = \operatorname{sm}(A)$.

Proof. Property (a) follows from characterization (b) in Theorem 2.13, properties (b) and (c) follow from the definition of the permanent, and property (d) follows from from characterization (d) in Theorem 2.13, the row-/column-wise symmetry of the scaling mean, and Proposition 2.5. $\qquad\square$

2.5. **Comparison between the two means.** We have the following inequalities between the two means we have defined:

Theorem 2.17 (Generalized van der Waerden bounds)**.** *For all $A \in \mathbb{R}_+^{n \times n}$,*

$$\operatorname{sm}(A) \leqslant \operatorname{pm}(A) \leqslant n(n!)^{-1/n} \operatorname{sm}(A). \tag{2.8}$$

Equality holds in the first inequality iff A has permanent 0 or rank 1. Equality holds in the second inequality iff A has permanent 0 or has the zero pattern of a permutation matrix.

The sequence $\big(n(n!)^{-1/n}\big)$ is increasing, and by Stirling's formula it converges to e; in particular:

$$\operatorname{sm}(A) \leqslant \operatorname{pm}(A) < e \operatorname{sm}(A) \quad \text{for all nonnegative square matrices } A. \tag{2.9}$$

Concerning the cases of equality in Theorem 2.17, we recall that by the Frobenius–König theorem [MM, Mi], a matrix in $\mathbb{R}_+^{n \times n}$ has zero permanent if and only if it contains a zero submatrix of size $r \times s$ with $r + s = n + 1$.

Notice that if $A \in \Omega_n$ then, by Proposition 2.4, inequalities (2.8) are equivalent to (2.1); in particular, the first inequality in (2.8) is a generalization of the van der Waerden bound to general nonnegative matrices. Actually, (2.8) is a corollary of (2.1), as we proceed to show.

Proof of Theorem 2.17. First consider $A \in \mathcal{P}_n$. By Theorem 2.15, A has a Sinkhorn decomposition DSE. The matrix S obeys the bounds (2.1) or equivalently (2.2), and so using Proposition 2.10 we obtain (2.8). The first inequality in (2.8) is an equality iff $A = D^{-1} J_n E^{-1}$, that is, iff A is a positive matrix of rank 1. The second inequality is an equality iff $A = D^{-1} P E^{-1}$ for some permutation matrix P, that is, iff A has the zero pattern of a permutation matrix.

Since $\mathcal{P}_n$ is dense in $\mathbb{R}_+^{n \times n}$, by continuity (Proposition 2.3) we conclude that inequalities (2.8) hold for every A.

The stated conditions for equality are clearly sufficient, so let us check necessity. Consider A such that $\operatorname{sm}(A) = \operatorname{pm}(A)$. By Proposition 2.16, $\operatorname{sm} \Pi(A) = \operatorname{pm} \Pi(A)$

and moreover there are two possibilities: either $\Pi(A) = 0$ or $\Pi(A) \in \mathcal{P}_n$. In the first case, A has permanent 0. In the second case, as we have seen above, $\Pi(A)$ is a positive matrix of rank 1; in particular $A = \Pi(A)$ has rank 1. This proves the characterization of the first equality. The second one is dealt with analogously. $\quad\square$

Let us review a few other results of the literature from our perspective. The paper [LSW] basically shows that the scaling mean can be computed efficiently, and therefore by (2.9) the permanental mean can be efficiently computed up to a factor of e, or equivalently, the permanent of a nonnegative $n \times n$ matrix can be efficiently computed up to a factor of e^n. Using instead the Bethe approximation for the permanent, it is possible to improve this factor to 2^n: see [GS]. We will see next further advantages of the scaling mean for the approximation of permanents.

2.6. **The scaling mean as a limit of permanental means.** Recall that the *Kronecker product* of two arbitrary matrices is defined as:

$$A \otimes B := \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{mr \times ns} \text{ if } A = (a_{ij}) \in \mathbb{R}^{m \times n}, \ B \in \mathbb{R}^{r \times s}.$$

It satisfies the following *mixed-product property* (see [MM, p. 9]):

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \tag{2.10}$$

Notice that the map Π considered in Proposition 2.16 has the following additional property:

$$\Pi(A \otimes B) = \Pi(A) \otimes B \quad \text{if } B \text{ is a } \textit{positive} \text{ square matrix.} \tag{2.11}$$

The scaling mean is well behaved with respect to the Kronecker product:

Proposition 2.18. *If A, B are nonnegative square matrices then*

$$\mathrm{sm}(A \otimes B) = \mathrm{sm}(A)\,\mathrm{sm}(B).$$

Proof. It is possible to prove the proposition directly from the definition of scaling mean, but let us give an alternative argument. By continuity, it is sufficient to consider matrices $A \in \mathbb{R}_+^{n \times n}$, $B \in \mathbb{R}_+^{m \times m}$ that have Sinkhorn decompositions, say $A = DSE$, $B = D'S'E'$. Then $A \otimes B$ also has a Sinkhorn decomposition, namely,

$$A \otimes B = (D \otimes D')(S \otimes S')(E \otimes E'),$$

so it follows from Proposition 2.10 that

$$\begin{aligned} \mathrm{sm}(A \otimes B) &= \frac{\mathrm{gm}(D \otimes D')\,\mathrm{gm}(E \otimes E')}{nm} \\ &= \frac{\mathrm{gm}(D)\,\mathrm{gm}(E)}{n} \cdot \frac{\mathrm{gm}(D')\,\mathrm{gm}(E')}{m} \\ &= \mathrm{sm}(A)\,\mathrm{sm}(B). \end{aligned} \qquad\square$$

The property expressed by Proposition 2.18 fails for permanental means: consider identity matrices, for example. On the other hand, Brualdi [Br] formulated a conjecture that can be restated in terms of permanental means as follows: If A, B are nonnegative square matrices then

$$\mathrm{pm}(A \otimes B) \leqslant \mathrm{pm}(A)\,\mathrm{pm}(B). \tag{2.12}$$

Based on experimental evidence, we also conjecture that equality holds iff A or B have permanent 0 or both A and B have rank 1. Unfortunately Brualdi's conjecture, formulated around 50 years ago, apparently has not received much attention.

We now describe an important relation between the permanental and the scaling means. Prior to the proof of the van der Waerden conjecture, Friedland [Fr] proved that the permanent of each $S \in \Omega_n$ is at least e^{-n}, which by the Stirling formula differs from the van der Waerden lower bound by a subexponential factor. The crux of his proof is to obtain the following limit formula:

$$\lim_{m \to \infty} (\operatorname{per}(S \otimes J_m))^{1/m} = e^{-n} \quad \text{if } S \in \Omega_n, \tag{2.13}$$

where J_m is the $m \times m$ matrix all of whose entries are $1/m$. We call this fact the *Friedland limit*; it appears in [Fr] as formula (1.6) and, as explained in § 2 of that paper, follows relatively easy from the then unproved van der Waerden bound. Basically as corollary of Friedland's limit, we will obtain the following:

Theorem 2.19 (Generalized Friedland limit). *For any nonnegative square matrix A we have*

$$\operatorname{sm}(A) = \lim_{m \to \infty} \operatorname{pm}(A \otimes U_m), \tag{2.14}$$

where U_m is the $m \times m$ matrix all of whose entries are 1.

Remark 2.20. The generalized van der Waerden bound (i.e., the first inequality in Theorem 2.17) becomes a corollary of Theorem 2.19 if Brualdi conjecture (2.12) is assumed.

Proof of Theorem 2.19. Consider $S \in \Omega_n$. By Stirling formula, $(m!)^{1/m} = e^{-1} m \theta_m$, where $\theta_m \to 1$, so

$$\operatorname{pm}(S \otimes U_m) = m \operatorname{pm}(S \otimes J_m) = m \left(\frac{\operatorname{per}(S \otimes J_m)}{(nm)!} \right)^{1/nm} = \frac{(\operatorname{per}(S \otimes J_m))^{1/nm}}{e^{-1} n \theta_{nm}},$$

which by Friedland's limit (2.13) converges to $1/n$. Since $\operatorname{sm}(S) = 1/n$, we conclude that (2.14) holds for doubly stochastic matrices. Next consider a matrix A in the set $\mathcal{P}_n$ (recall Definition 2.14), and let DSE be its Sinkhorn decomposition. Using the mixed-product property (2.10) we factorize $A \otimes U_m$ as $(D \otimes I_m)(S \otimes U_m)(E \otimes I_m)$, where the matrices $D \otimes I_m$ and $E \otimes I_m$ are diagonal. Therefore

$$\operatorname{pm}(A \otimes U_m) = \underbrace{\operatorname{gm}(D \otimes I_m)}_{=\operatorname{gm}(D)} \underbrace{\operatorname{pm}(S \otimes U_m)}_{\to 1/m} \underbrace{\operatorname{gm}(E \otimes I_m)}_{=\operatorname{gm}(E)} \to \operatorname{sm}(A),$$

establishing (2.14) in this case. Finally, consider a general $A \in \mathbb{R}_+^{n \times n}$. Using Proposition 2.16 and property (2.11) we obtain

$$\operatorname{pm}(A \otimes U_m) = \operatorname{pm}(\Pi(A \otimes U_m)) = \operatorname{pm}(\Pi(A) \otimes U_m) \to \operatorname{sm}(\Pi(A)) = \operatorname{sm}(A). \quad \square$$

As mentioned in the end of § 2.5, the scaling mean is computationally easier to compute than the permanental mean, and in view of the bounds of Theorem 2.17, the former can be used to approximate the latter up to a multiplicative error of at most e. Theorem 2.19 indicates that for large matrices of the form $A \otimes U_m$, this factor is actually close to 1.

As we will see later, some classes of random matrices are approximately permutationally equivalent to matrices of the form $A \otimes U_m$: this is the idea behind the Law of Large Permanents. While in this paper we are not concerned with numerical

studies, it may be interesting to investigate further the effectiveness of the scaling mean for the computation of permanents of reasonably well-behaved matrices.

Remark 2.21. In addition to (2.3), (2.4), (2.7), and (2.14), other characterizations of the scaling mean are [FLS, Thrm. 5.3], [GS, formula (2)] and the following one:

$$\mathrm{sm}(A) = \frac{1}{n} \inf_{\Delta} \frac{\rho(\Delta A)}{\mathrm{gm}(\Delta)} \,, \tag{2.15}$$

where Δ runs over the diagonal matrices with positive main diagonal. We only sketch the proof: The $\leqslant$ inequality follows from (2.6) and row-wise homogeneity of the scaling mean. In the case that $A \in \mathcal{P}_n$, i.e. A has a Sinkhorn decomposition DSE, the infimum on the RHS is attained at $\Delta = E^{-1}D^{-1}$. Recall that the LHS of (2.15) depends continuously on A, while the RHS is upper semicontinuous and monotonically increasing with respect to matrix entries. (The latter fact follows from monotonicity of ρ, itself a consequence of Gelfand spectral radius formula [Bh, p. 204].) This permits us to extend the equality from the dense set $\mathcal{P}_n$ to the whole $\mathbb{R}_+^{n \times n}$.

3. Scaling mean and Sinkhorn decomposition of functions

In this section we extend the notions of scaling mean and Sinkhorn decomposition from matrices to arbitrary nonnegative measurable functions defined on probability spaces with respect to a pair of sub-σ-algebras. The allowable scalings are those that are measurable with respect to one of these sub-σ-algebras, which thus supersede the row/column structure.

Moreover we prove the existence of this functional Sinkhorn decomposition under a boundedness assumption (Theorem 3.6).

Finally, in § 3.5 we analyze the particular setting of direct products, which is the one that we will use later for our Law of Large Permanents (Section 4) and for its applications (Section 5). Let us remark that though this particular setting is sufficient for the main results of this paper, the general setting where scaling are controlled by sub-σ-algebras is in our opinion more transparent and, as discussed in Section 6, should be indispensable for stronger Laws.

3.1. **The functional scaling mean.** Let us fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\mathcal{G}(\mathbb{P})$ denote the set of positive measurable functions $\varphi \colon \Omega \to \mathbb{R}_{++}$ such that $\log \varphi \in L^1(\mathbb{P})$. The *geometric mean* of a function $f \in \mathcal{G}(\mathbb{P})$ is defined as:

$$\mathrm{gm}(\varphi) \coloneqq \exp\left(\textstyle\int \log \varphi \, \mathrm{d}\mathbb{P}\right) . \tag{3.1}$$

The AM-GM inequality says that $\mathrm{gm}(\varphi) \leqslant \int \varphi \, \mathrm{d}\mathbb{P}$.

Let us also fix a pair of sub-σ-algebras $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$. For each $i \in \{1, 2\}$, we define two sets of functions $\mathcal{G}_i \supset \mathcal{B}_i$ as follows:

$$\mathcal{G}_i \coloneqq \left\{\varphi \colon \Omega \to \mathbb{R}_{++} \, : \, \varphi \text{ is } \mathcal{A}_i\text{-measurable and } \log \varphi \in L^1(\mathbb{P})\right\}, \tag{3.2}$$

$$\mathcal{B}_i \coloneqq \left\{\varphi \colon \Omega \to \mathbb{R}_{++} \, : \, \varphi \text{ is } \mathcal{A}_i\text{-measurable and } \log \varphi \in L^\infty(\mathbb{P})\right\}. \tag{3.3}$$

Our scaling functions will always take values in these spaces.

Definition 3.1. Let $f \colon \Omega \to \mathbb{R}_+$ be a nonnegative measurable function. The *scaling mean* of f with respect to the sub-σ-algebras $\mathcal{A}_1, \mathcal{A}_2$ is defined as:

$$\mathrm{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f) \coloneqq \inf_{\substack{g_1 \in \mathcal{G}_1 \\ g_2 \in \mathcal{G}_2}} \frac{1}{\mathrm{gm}(g_1)\,\mathrm{gm}(g_2)} \int g_1 f g_2 \, \mathrm{d}\mathbb{P} \,. \tag{3.4}$$

When no confusion is likely to arise, we write $\mathrm{sm}(f) = \mathrm{sm}_{\mathcal{A}_1,\mathcal{A}_2}(f)$.

Concrete examples will be presented later in § 3.5.

We list some basic properties of the scaling mean:

- *Monotonicity:* If $f \leqslant g$ almost everywhere then $\mathrm{sm}(f) \leqslant \mathrm{sm}(g)$.
- *Reflexivity:* If f equals a constant c almost everywhere then $\mathrm{sm}(f) = c$; this is a consequence of the AM-GM inequality.
- *Homogeneity:* If $g_1 \in \mathcal{G}_1$ and $g_2 \in \mathcal{G}_2$ then

$$\mathrm{sm}(g_1 f g_2) = \mathrm{gm}(g_1)\,\mathrm{sm}(f)\,\mathrm{gm}(g_2)\,. \tag{3.5}$$

The following proposition says that in order to evaluate the infimum in formula (3.4) it is sufficient to consider g_i in the smaller space $\mathcal{B}_i$ defined by (3.3):

Proposition 3.2. *For any measurable $f \colon \Omega \to \mathbb{R}_+$ we have:*

$$\mathrm{sm}(f) := \inf_{\substack{g_1 \in \mathcal{B}_1 \\ g_2 \in \mathcal{B}_2}} \frac{1}{\mathrm{gm}(g_1)\,\mathrm{gm}(g_2)} \int g_1 f g_2 \,\mathrm{d}\mathbb{P}\,.$$

Proof. Fix $f \geqslant 0$ and consider $g_1 \in \mathcal{G}_1$, $g_2 \in \mathcal{G}_2$ such that $\int g_1 f g_2 \,\mathrm{d}\mathbb{P} < \infty$. Define two sequences $(g_{1,k})$, $(g_{2,k})$ respectively in $\mathcal{B}_1$, $\mathcal{B}_2$ as follows:

$$g_{i,k}(\omega) := \begin{cases} g_i(\omega) & \text{if } |\log g_1(\omega)| \leqslant k, \\ 1 & \text{otherwise.} \end{cases}$$

Applying the dominated convergence theorem three times, we conclude that:

$$\frac{1}{\mathrm{gm}(g_1)\,\mathrm{gm}(g_2)} \int g_1 f g_2 \,\mathrm{d}\mathbb{P} = \lim_{k \to \infty} \frac{1}{\mathrm{gm}(g_{1,k})\,\mathrm{gm}(g_{2,k})} \int g_{1,k} f g_{2,k} \,\mathrm{d}\mathbb{P}\,.$$

The proposition follows. $\qquad\square$

3.2. **Conditional expectations and doubly stochastic functions.** We begin by recalling some basic facts about conditional expectations, referring the reader to [Bo] for more details.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. If $f \in L^1(\mathbb{P})$ and $\mathcal{A}_1$ is a sub-σ-algebra of $\mathcal{A}$, let $\mathbb{E}(f|\mathcal{A}_1)$ denote the *conditional expectation* of f with respect to $\mathcal{A}_1$, i.e. the a.e.-unique $\mathcal{A}_1$-measurable function such that

$$\int g f \,\mathrm{d}\mathbb{P} = \int g\,\mathbb{E}(f|\mathcal{A}_1) \,\mathrm{d}\mathbb{P} \quad \text{for every bounded } \mathcal{A}_1\text{-measurable function } g.$$

For example, if $\{B_1, \dots B_k\}$ is a partition of Ω into finitely many sets of positive probability and $\mathcal{A}_1$ is the σ-algebra generated by this partition, then $\mathbb{E}(f|\mathcal{A}_1)$ is the simple function that takes the value $\frac{1}{\mathbb{P}(B_i)} \int_{B_i} f \,\mathrm{d}\mathbb{P}$ on the set B_i.

Note that if $f \in L^1(\mathbb{P})$ then, as an immediate consequence of the definition of conditional expectation,

$$\mathbb{E}(g f|\mathcal{A}_1) = g\,\mathbb{E}(f|\mathcal{A}_1) \quad \text{for every bounded } \mathcal{A}_1\text{-measurable function } g. \tag{3.6}$$

Again fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a pair of sub-σ-algebras $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$. Let us say that a integrable nonnegative function $f \colon \Omega \to \mathbb{R}_+$ is *doubly stochastic* if

$$\mathbb{E}(f|\mathcal{A}_1) = \mathbb{E}(f|\mathcal{A}_2) = 1 \quad \mathbb{P}\text{-a.e.} \tag{3.7}$$

Notice that a doubly stochastic function has arithmetic mean $\int f \,\mathrm{d}\mathbb{P} = 1$. Analogously for the scaling mean:

Proposition 3.3. *If f is doubly stochastic then* $\mathrm{sm}(f) = 1$.

Proof. If f is doubly stochastic then the measure $\mathrm{d}\nu := f\,\mathrm{d}\mathbb{P}$ is a probability. Therefore, for any $g_1 \in \mathcal{B}_1$ and $g_2 \in \mathcal{B}_2$, using the AM-GM inequality and property (3.6) we obtain:

$$
\begin{aligned}
\int g_1 f g_2 \,\mathrm{d}\mathbb{P} &= \int g_1 g_2 \,\mathrm{d}\nu \\
&\geqslant \exp\left(\int \log(g_1 g_2)\,\mathrm{d}\nu\right) \\
&= \exp\left(\int f \log g_1 \,\mathrm{d}\mathbb{P}\right) \exp\left(\int f \log g_2 \,\mathrm{d}\mathbb{P}\right) \\
&= \exp\left(\int \mathbb{E}(f \log g_1 | \mathcal{A}_1)\,\mathrm{d}\mathbb{P}\right) \exp\left(\int \mathbb{E}(f \log g_2 | \mathcal{A}_2)\,\mathrm{d}\mathbb{P}\right) \\
&= \exp\left(\int (\log g_1)\mathbb{E}(f|\mathcal{A}_1)\,\mathrm{d}\mathbb{P}\right) \exp\left(\int (\log g_2)\mathbb{E}(f|\mathcal{A}_2)\,\mathrm{d}\mathbb{P}\right) \\
&= \mathrm{gm}(g_1)\,\mathrm{gm}(g_2).
\end{aligned}
$$

So it follows from Proposition 3.2 that $\mathrm{sm}(f) \geqslant 1$. Considering $g_1 = g_2$ identically equal to 1 we conclude that $\mathrm{sm}(f) = 1$. $\qquad\square$

3.3. **Functional Sinkhorn decompositions.**

Fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a pair of sub-σ-algebras $\mathcal{A}_1$, $\mathcal{A}_2 \subset \mathcal{A}$.

Definition 3.4. A *Sinkhorn decomposition* of a function $f\colon \Omega \to \mathbb{R}_+$ is a factorization of the form

$$
f(\omega) = \varphi(\omega) g(\omega) \psi(\omega)
$$

where g is doubly stochastic, $\varphi \in \mathcal{G}_1$, and $\psi \in \mathcal{G}_2$.

Again we postpone the examples to § 3.5.

Let us relate the scaling mean and Sinkhorn decompositions:

Proposition 3.5. *If f has a Sinkhorn decomposition $\varphi g \psi$ then*

    (a) $\mathrm{sm}(f) = \mathrm{gm}(\varphi)\,\mathrm{gm}(\psi)$.
    (b) *The infimum in* (3.4) *is attained at the functions* $g_1 = 1/\varphi$, $g_2 = 1/\psi$.

Proof. Part (a) is an immediate consequence of Proposition 3.3 and the homogeneity property (3.5). Part (b) follows from part (a). $\qquad\square$

Our next result gives a sufficient condition for the existence of Sinkhorn decompositions. We denote by $\mathcal{B}(\mathbb{P})$ the space of positive functions f such that $\log f \in L^\infty(\mathbb{P})$.

Theorem 3.6. *Every $f \in \mathcal{B}(\mathbb{P})$ has a Sinkhorn decomposition $\varphi g \psi$ where $\varphi \in \mathcal{B}_1$ and $\psi \in \mathcal{B}_2$. Conversely, if $f = \varphi' g' \varphi'$ is another Sinkhorn decomposition such that $\varphi' \in \mathcal{B}_1$ and $\psi' \in \mathcal{B}_2$ then there exists $\lambda \in \mathbb{R}_{++}$ such that $\varphi' = \lambda\varphi$, $g' = g$, and $\psi' = \lambda^{-1}\psi$ $\mathbb{P}$-almost everywhere.*

The proof of this theorem is independent of the rest of the paper and is presented in the next subsection. The theorem itself (in the particular case of directed products: see § 3.5) will be used in the proof of our Law of Large Permanents (Theorem 4.1) in Section 4.

3.4. **Proof of Theorem 3.6.** The proof has two steps. First, we reduce the problem to the existence of a fixed point for a certain nonlinear operator. Then we show that this operator contracts Hilbert's projective metric, and so it must have a fixed point. These ideas come from the literature. The fact that the existence of Sinkhorn decompositions for matrices is equivalent to a fixed point problem was first noted by Menon [Me1]. The usefulness of Hilbert's projective metric in this context was noted independently in [FL] (for matrices) and [Nu1] (for matrices and for functions). Unfortunately this elegant strategy uses strict positivity in an essential way, and in order to study more general functions other methods are needed: see [Nu2, BLN].

Fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a pair of sub-$\sigma$-algebras $\mathcal{A}_1$, $\mathcal{A}_2 \subset \mathcal{A}$, and a function $f \in \mathcal{B}(\mathbb{P})$. Let $\mathcal{B}_1$, $\mathcal{B}_2 \subset \mathcal{B}(\mathbb{P})$ be as in (3.3). Define four maps forming a (non-commutative) diagram:

$$
\begin{array}{ccc}
\mathcal{B}_1 & \xrightarrow{\ \mathtt{I}_1\ } & \mathcal{B}_1 \\[4pt]
{\scriptstyle \mathtt{K}_1}\big\uparrow & & \big\downarrow{\scriptstyle \mathtt{K}_2} \\[4pt]
\mathcal{B}_2 & \xleftarrow[\ \mathtt{I}_2\ ]{} & \mathcal{B}_2
\end{array}
$$

by the formulas:

$$
(\mathtt{I}_i(\varphi))(\omega) := \frac{1}{\varphi(\omega)}, \qquad (\mathtt{K}_i(\varphi))(\omega) := \mathbb{E}(\varphi f | \mathcal{A}_i)(\omega).
$$

Let $\mathtt{T} := \mathtt{K}_1 \circ \mathtt{I}_2 \circ \mathtt{K}_2 \circ \mathtt{I}_1$, i.e., the map obtained by going around the diagram. Notice that $\mathtt{T}$ maps rays (i.e. sets of the form $\mathbb{R}_{++}\varphi$) into rays. The following observation says that fixed rays yield Sinkhorn decompositions:

Lemma 3.7. *Suppose $\varphi \in \mathcal{B}_1$ and $c \in \mathbb{R}_{++}$ are such that $\mathtt{T}(\varphi) = c\varphi$ $\mathbb{P}$-a.e. Then $c = 1$ and moreover, letting $\psi := \mathtt{K}_2 \circ \mathtt{I}_1(\varphi)$ and $g := f/(\varphi\psi)$, the factorization $\varphi g \psi$ is a Sinkhorn decomposition of f. Conversely, if $\varphi g \psi$ is a Sinkhorn decomposition of f with $\varphi \in \mathcal{B}_1$ and $\psi \in \mathcal{B}_2$ then $\mathtt{T}(\varphi) = \varphi$ $\mathbb{P}$-a.e.*

Proof. Suppose $\mathtt{T}(\varphi) = c\varphi$ $\mathbb{P}$-a.e., and let ψ and g be as above. Then the following equalities hold $\mathbb{P}$-a.e.:

$$
\begin{aligned}
\mathbb{E}(g | \mathcal{A}_2) &= \mathbb{E}\big(\tfrac{f}{\varphi\psi} \big| \mathcal{A}_2\big) && \text{(by definition of } g) \\
&= \tfrac{1}{\psi}\mathbb{E}\big(\tfrac{f}{\varphi} \big| \mathcal{A}_2\big) && \text{(by property (3.6) and the fact that } \psi \in \mathcal{B}_2) \\
&= 1 && \text{(by definition of } \psi).
\end{aligned}
$$

Similarly, the relation $\mathtt{K}_1 \circ \mathtt{I}_2(\psi) = c\varphi$ implies that

$$
\mathbb{E}(g | \mathcal{A}_1) = c \quad \mathbb{P}\text{-a.e.}
$$

Integrating with respect to $\mathbb{P}$ yields that $c = 1$, and therefore g is doubly stochastic, as we wanted to show. The converse part of the lemma is immediate. $\qquad\square$

Hence to complete the proof of Theorem 3.6 we are left to show that $\mathtt{T}$ has a fixed ray. We will use a classical geometric–analytical device.

Given two functions φ, $\hat{\varphi} \in \mathcal{B}_1$, we define their *Hilbert distance* as:

$$
d(\varphi, \hat{\varphi}) := \log\left(\frac{\operatorname{ess\,sup}(\hat{\varphi}/\varphi)}{\operatorname{ess\,inf}(\hat{\varphi}/\varphi)}\right). \tag{3.8}
$$

This is a pseudometric such that $d(\varphi, \hat{\varphi}) = 0$ iff $\hat{\varphi} = c\varphi$ $\mathbb{P}$-a.e. for some constant $c \in \mathbb{R}_{++}$. In other words, if elements of $\mathcal{B}_1$ that coincide a.e. are considered equal then d induces a genuine metric on the quotient space $\mathcal{B}_1/\mathbb{R}_{++}$. Moreover, this metric is complete. Analogously, we define a pseudometric on $\mathcal{B}_2$, which we also denote by d. Definition (3.8) is only a particular case of Hilbert's projective metric on convex cones; see e.g. [Nu1] or [Li, Section 1]. In these references the reader will find an important property due to Garrett Birkhoff [Bi], which specialized to our case is stated as follows:

Proposition 3.8. *If* $\mathtt{L}\colon \mathcal{B}_1 \to \mathcal{B}_2$ *is linear and its image has finite diameter δ then the following contraction property holds:*

$$d(\mathtt{L}(\varphi), \mathtt{L}(\hat{\varphi})) \leqslant \left(\tanh \frac{\delta}{4} \right) d(\varphi, \hat{\varphi}) \quad \text{for all } \varphi, \hat{\varphi} \in \mathcal{B}_1 \,.$$

Proof of Theorem 3.6. The linear map $\mathtt{K}_2(\varphi) := \mathbb{E}(\varphi f | \mathcal{A}_2)$ defined before satisfies the inequalities

$$(\operatorname{ess\,inf} f)\, \mathbb{E}(\varphi | \mathcal{A}_2) \leqslant \mathtt{K}_2(\varphi) \leqslant (\operatorname{ess\,sup} f)\, \mathbb{E}(\varphi | \mathcal{A}_2) \quad \mathbb{P}\text{-a.e.}$$

Analogous inequalities for $\hat{\varphi}$ yield

$$\left(\frac{\operatorname{ess\,inf} f}{\operatorname{ess\,sup} f} \right) \frac{\mathbb{E}(\varphi | \mathcal{A}_2)}{\mathbb{E}(\hat{\varphi} | \mathcal{A}_2)} \leqslant \frac{\mathtt{K}_2(\varphi)}{\mathtt{K}_2(\hat{\varphi})} \leqslant \left(\frac{\operatorname{ess\,sup} f}{\operatorname{ess\,inf} f} \right) \frac{\mathbb{E}(\varphi | \mathcal{A}_2)}{\mathbb{E}(\hat{\varphi} | \mathcal{A}_2)} \,,$$

which by (3.8) imply that the image of $\mathtt{K}_2$ has diameter

$$\delta \leqslant 2 \log \left(\frac{\operatorname{ess\,sup} f}{\operatorname{ess\,inf} f} \right) < \infty \,.$$

So by the Proposition 3.8 the map $\mathtt{K}_2$ contracts Hilbert distances uniformly. Analogously for $\mathtt{K}_1$. On the other hand, the nonlinear maps $\mathtt{I}_1$ and $\mathtt{I}_2$ preserve Hilbert distances. In particular $\mathtt{T} = \mathtt{K}_1 \circ \mathtt{I}_2 \circ \mathtt{K}_2 \circ \mathtt{I}_1$ contracts Hilbert distances uniformly. By completeness we conclude that $\mathtt{T}$ has an unique fixed ray, so Lemma 3.7 allows us to conclude. $\qquad\square$

As a byproduct of the proof above, the iterates under $\mathtt{T}$ of any initial ray converge with a known exponential rate to a fixed ray. This gives an effective algorithm for the computation of a Sinkhorn decomposition of f and of its scaling mean, in particular.

3.5. **The main example: direct products.** Here we describe the particular form of the previous results that will be used in Sections 4 and 5.

Let $(X, \mathcal{X}, \mu), (Y, \mathcal{Y}, \nu)$ be two probability spaces. Take $(\Omega, \mathcal{A}, \mathbb{P})$ as the product space $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu)$, and take the two sub σ-algebras on $\mathcal{A}_1 = \mathcal{X} \times \{\varnothing, Y\}$, $\mathcal{A}_2 = \{\varnothing, X\} \times \mathcal{Y}$.

An $\mathcal{A}_1$-measurable function is just a function that only depends on the x coordinate, and analogously for $\mathcal{A}_2$-measurable functions. We denote by $\mathcal{G}(\mu)$, respectively $\mathcal{B}(\mu)$, the sets of positive measurable functions $\varphi\colon X \to \mathbb{R}_{++}$ such that $\log \varphi$ is μ-integrable, respectively, μ-essentially bounded. Analogously we define sets $\mathcal{G}(\nu)$ and $\mathcal{B}(\nu)$. Recalling notations (3.2), (3.3), with a small abuse of language we can write

$$\mathcal{G}_1 = \mathcal{G}(\mu), \quad \mathcal{B}_1 = \mathcal{B}(\mu), \quad \mathcal{G}_2 = \mathcal{G}(\nu), \quad \mathcal{B}_2 = \mathcal{B}(\nu).$$

In this context, the previously introduced concepts can be recast as follows:

- The scaling mean of a measurable function $f \colon X \times Y \to \mathbb{R}_+$ is:

$$\operatorname{sm}(f) = \inf_{\substack{\varphi \in \mathcal{G}(\mu) \\ \psi \in \mathcal{G}(\nu)}} \frac{1}{\operatorname{gm}(\varphi) \operatorname{gm}(\psi)} \iint \varphi(x) f(x,y) \psi(y) \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) \,. \qquad (3.9)$$

- The conditional expectations of a measurable function $f \colon X \times Y \to \mathbb{R}$ are:

$$\mathbb{E}(f|\mathcal{A}_1) = \smallint f(\cdot, y) \, \mathrm{d}\nu(y) \quad \text{and} \quad \mathbb{E}(f|\mathcal{A}_2) = \smallint f(x, \cdot) \, \mathrm{d}\mu(x).$$

- A function $f \colon X \times Y \to \mathbb{R}_+$ is doubly stochastic if:

$$\smallint f(\cdot, y) \, \mathrm{d}\mu = \smallint f(x, \cdot) \, \mathrm{d}\nu = 1 \quad \text{for } \mu\text{-a.e. } x \text{ and } \nu\text{-a.e. } y. \qquad (3.10)$$

- A Sinkhorn decomposition of a function $f \colon X \times Y \to \mathbb{R}_+$ is a factorization of the form:

$$f(x,y) = \varphi(x) g(x,y) \psi(y) \qquad (3.11)$$

where g is doubly stochastic, $\varphi \in \mathcal{G}(\mu)$, and $\psi \in \mathcal{G}(\nu)$.

In the case that X and Y are finite sets and the probabilities μ and ν are equidistributed, we reobtain the concepts studied in Section 2; notice however that the definition of doubly stochastic matrices uses a different normalization.

Let us remark that functional Sinkhorn decompositions as in (3.11) were first obtained by Knopp and Sinkhorn [KS] assuming that X and Y are compact and that f is continuous and strictly positive. For other existence results under various hypotheses, see [Nu2, BLN].

Let us see a concrete non-trivial situation where functional Sinkhorn decompositions and scaling means can be computed directly. The result below will be used in Section 5.

Proposition 3.9. *Suppose that $f \colon X \times Y \to \mathbb{R}_+$ is a measurable nonnegative function of the form*

$$f(x,y) = \begin{cases} f_0(x) & \text{if } y \in Y_0 \,, \\ f_1(x) & \text{if } y \in Y_1 \,, \end{cases}$$

where $\{Y_0, Y_1\}$ is a partition of Y into sets of positive measure, and $\log(f_0 + f_1)$ is μ-integrable. Then

$$\operatorname{sm}(f) = c^c \left(\frac{1-c}{r} \right)^{1-c} \exp \left(\int \log(f_0 + r f_1) \, \mathrm{d}\mu \right), \qquad (3.12)$$

where $c := \nu(Y_0)$ and r is the unique positive root of the equation

$$\int \frac{f_0}{f_0 + r f_1} \, \mathrm{d}\mu = c \,. \qquad (3.13)$$

Proof. By dominated convergence, the LHS of (3.13) is continuous and increasing with respect to r, and so by the intermediate value theorem the equation has a unique solution $r \in (0, \infty)$. Let $\varphi := f_0 + r f_1$. Note that $|\log \varphi - \log(f_0 + f_1)| \leqslant |\log r|$, and therefore $\varphi \in \mathcal{G}(\mu)$. Let

$$\psi(y) := \begin{cases} c & \text{if } y \in Y_0 \,, \\ (1-c)/r & \text{if } y \in Y_1 \,. \end{cases}$$

Direct calculation shows that the function $g := f/(\varphi\psi)$ is doubly stochastic, so $\varphi g \psi$ is a Sinkhorn decomposition of f. Proposition 3.5(a) then yields (3.12). $\qquad \square$

4. A Law of Large Permanents

4.1. **Statement and comments.** In this section we prove our main result (Theorem 4.1), which was already stated in the Introduction. Let us recall the statement for the reader's convenience, and also fix some notation.

Assume that $(X, \mathcal{X}, \mu)$, $(Y, \mathcal{Y}, \nu)$ are Lebesgue probability spaces, and $T\colon X \to X$, $S\colon Y \to Y$ are measure preserving transformations. Given a function $f\colon X \times Y \to \mathbb{R}$, for each $(x, y) \in X \times Y$ and each positive integer n we define the following $n \times n$ matrix:

$$\square_n f(x, y) := \begin{pmatrix} f(x, y) & f(Tx, y) & \cdots & f(T^{n-1}x, y) \\ f(x, Sy) & f(Tx, Sy) & \cdots & f(T^{n-1}x, Sy) \\ \vdots & \vdots & & \vdots \\ f(x, S^{n-1}y) & f(Tx, S^{n-1}y) & \cdots & f(T^{n-1}x, S^{n-1}y) \end{pmatrix}, \qquad (4.1)$$

which can be thought as the truncation of an infinite matrix $\square f(x, y) \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$.

Recall that $\mathcal{B}(\mu \times \nu)$ denotes the set of positive measurable functions on $X \times Y$ essentially bounded away from zero and infinity. Such functions have scaling means, given by formula (3.9). We can now restate our main result as follows:

Theorem 4.1. *If T and S are ergodic, and $f \in \mathcal{B}(\mu \times \nu)$ then*

$$\lim_{n \to \infty} \mathrm{pm}\left(\square_n f(x, y)\right) = \mathrm{sm}(f) \qquad (4.2)$$

for $\mu \times \nu$-almost every $(x, y) \in X \times Y$.

This *Law of Large Permanents* is a very general ergodic theorem. In Section 5 we will see how to apply it to other natural types of means, and in Section 6 we will discuss the possibility of even more general laws. We stress that Theorem 4.1 not only states the existence of the limit in (4.2), but characterizes it as a scaling mean, which can be used to efficiently compute its value.

It is worthwhile to note that Theorem 4.1 implies the generalized Friedland limit (Theorem 2.19), at least for positive matrices. Indeed, given $A \in \mathbb{R}_{++}^{k \times k}$, take T and S as cyclic permutations on sets X and Y of cardinality k, and let μ and ν be the corresponding invariant measures. Let f be such that $\square_k f(x, y) = A$ for some point (x, y). Then, for every $m \geqslant 1$, the matrix $\square_{km} f(x, y)$ is permutationally equivalent to the Kronecker product $A \otimes U_m$, where U_m is the $m \times m$ matrix all of whose entries are 1. In particular, $\mathrm{pm}(A \otimes U_m)$ equals $\mathrm{pm}(\square_{km} f(x, y))$, which by Theorem 4.1 converges to $\mathrm{sm}(f)$ as $k \to \infty$. Using for instance a Sinkhorn decomposition of the matrix A and the related Sinkhorn decomposition of the function f, one checks that $\mathrm{sm}(f) = \mathrm{sm}(A)$, thus obtaining the generalized Friedland limit (2.14).

Actually, we will reason in the converse direction and deduce Theorem 4.1 from Theorem 2.19. Let us sketch the proof. The first step is to approximate f by a suitable simple function, and to show that the permanental means do not change much; the values of the simple function are recorded on a square matrix A. The second step is to show that the matrix (4.1) is, modulo a permutation of rows and columns, approximately equal to a Kronecker product $A \otimes U_m$, and then to use Theorem 2.19 to relate the permanental means of these matrices with scaling means. It is technically convenient to work with doubly stochastic functions, so we will also use Theorem 3.6 on the existence of Sinkhorn decompositions.

The precise proof of Theorem 4.1 will take up the rest of this section, which is organized as follows: In § 4.2 we recall some basic results: approximation by

conditional expectations, and an ergodic theorem. In § 4.3 we obtain some estimates on how the permanent changes under perturbations of the matrix. These facts are used in § 4.4 to implement the strategy sketched above and to prove the theorem.

4.2. **Preliminaries from Measure Theory and Ergodic Theory.** We have defined conditional expectations in § 3.2. The following result is contained in [Bo, Theorems 10.2.2, 10.2.3] and describes the behavior of conditional expectations as the σ-algebra is refined:

Proposition 4.2. *Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and let $f \in L^1(\mu)$. Suppose that $(\mathcal{A}_k)$ is an increasing sequence of sub-σ-algebras of $\mathcal{A}$ whose union generates the σ-algebra $\mathcal{A}$ modulo sets of measure zero. Then the functions $\mathbb{E}(f|\mathcal{A}_k)$ converge almost surely and in L^1 to f as $k \to \infty$.*

Let us describe the basic ergodic theorem that we will need. Consider probability spaces (X, μ), (Y, ν) and measure preserving transformations $T\colon X \to X$, $S\colon Y \to Y$. Then we can define a measure preserving action of the semigroup $\mathbb{N}^2$ (where $0 \in \mathbb{N}$) on the product space $(X \times Y, \mu \times \nu)$ as follows:

$$\mathcal{T}^{(i,j)}(x,y) := (T^i x, S^j y), \quad \text{where } (i,j) \in \mathbb{N}^2 \text{ and } (x,y) \in X \times Y.$$

Notice that this action is ergodic if and only if both T and S are ergodic. In this case, by the ergodic theorem for $\mathbb{N}^2$-actions (see [Ke, Theorem 2.1.5] or [Kr, Chapter 6, Theorem 3.5]), for any $h \in L^1(\mu \times \nu)$ we have

$$\lim_{n \to \infty} \frac{1}{n^2} \sum_{i,j=0}^{n-1} h(T^i x, S^j y) = \int h \, \mathrm{d}(\mu \times \nu) \qquad \text{for } \mu \times \nu\text{-a.e. } (x,y). \tag{4.3}$$

4.3. **More preliminaries: Regularity estimates for the permanent of strictly positive matrices.** Let us begin by recalling some basic facts. As for the determinant, the permanent of a square matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ can be computed by means of a *Laplace expansion* along any column j:

$$\operatorname{per} A = \sum_{i=1}^{n} a_{ij} \operatorname{per} A(i|j),$$

where, as usual, $A(i|j)$ denotes the matrix obtained from A by deleting the i-th row and the j-th column. Similar Laplace expansions along rows hold. Using either kind of expansion, we see that the partial derivatives of the permanent function are simply:

$$\frac{\partial \operatorname{per} A}{\partial a_{ij}} = \operatorname{per} A(i|j). \tag{4.4}$$

Given $\lambda > 1$, let us say that a positive matrix $A = (a_{ij}) \in \mathbb{R}_{++}^{n \times n}$ is λ-*bounded* if:

$$\lambda^{-1} \leqslant a_{ij} \leqslant \lambda \quad \text{for each } i, j. \tag{4.5}$$

Lemma 4.3. *Let $A \in \mathbb{R}_{++}^{n \times n}$ be a λ-bounded matrix and denote by s be the sum of the entries of A. Then for any $i, j \in \{1, \ldots, n\}$ we have*

$$\lambda^{-5} n \operatorname{per} A \leqslant \lambda^{-4} n s^{-1} \operatorname{per} A \leqslant \operatorname{per} A(i|j) \leqslant \lambda^4 n s^{-1} \operatorname{per} A \leqslant \lambda^5 n \operatorname{per} A \, .$$

Proof. The outer inequalities being trivial, we only need to care about the inner ones. Summing Laplace expansions of the permanent of $A = (a_{ij})$ along the n columns, we obtain $n \operatorname{per} A = \sum_{i,j} a_{ij} \operatorname{per} A(i|j)$, and in particular,

$$s \min_{i,j} \operatorname{per} A(i|j) \leqslant n \operatorname{per} A \leqslant s \max_{i,j} \operatorname{per} A(i|j) \,.$$

Suppose that the minimum and the maximum above are attained at (i_1, j_1) and (i_2, j_2), respectively. Consider the two Laplace expansions:

$$\operatorname{per} A(i_2|j_2) = \sum_{k \neq i_2} a_{kj_1} \operatorname{per} A(i_2 k|j_1 j_2) \,,$$

$$\operatorname{per} A(i_2|j_1) = \sum_{k \neq i_2} a_{kj_2} \operatorname{per} A(i_2 k|j_1 j_2) \,.$$

Since $a_{kj_1} \leqslant \lambda^2 a_{kj_2}$, it follows that $A(i_2|j_2) \leqslant \lambda^2 \operatorname{per} A(i_2|j_1)$. An analogous argument gives $\operatorname{per} A(i_2|j_1) \leqslant \lambda^2 \operatorname{per} A(i_1|j_1)$, thus proving that $\operatorname{per} A(i_2|j_2) \leqslant \lambda^4 \operatorname{per} A(i_1|j_1)$. So for any i, j we have

$$\operatorname{per} A(i|j) \leqslant \operatorname{per} A(i_2|j_2) \leqslant \lambda^4 \operatorname{per} A(i_1|j_1) \leqslant \lambda^4 n s^{-1} \operatorname{per} A \,,$$

as claimed. The remaining inequality is proved analogously. $\qquad \square$

We can now prove a regularity estimate for the permanent:

Lemma 4.4. *If A, $B \in \mathbb{R}_{++}^{n \times n}$ are λ-bounded matrices then*

$$\left| \log \frac{\operatorname{per} B}{\operatorname{per} A} \right| \leqslant \frac{\lambda^4 n \sum_{i,j} |b_{ij} - a_{ij}|}{\min \left\{ \sum_{i,j} a_{ij}, \ \sum_{i,j} b_{ij} \right\}} \leqslant \frac{\lambda^5}{n} \sum_{i,j} |b_{ij} - a_{ij}| \,.$$

Proof. The second inequality being trivial, we only need to care about the first one. For all $t \in [0,1]$, the convex combination $A_t := (1-t)A + tB$ is also a positive λ-bounded matrix. To prove the first inequality, we will apply the mean value theorem to the function $f(t) := \log \operatorname{per} A_t$ and use the estimate

$$\left| \log \frac{\operatorname{per} B}{\operatorname{per} A} \right| = |f(1) - f(0)| \leqslant \max_{t \in [0,1]} |f'(t)| \,. \tag{4.6}$$

Using formula (4.4), we compute:

$$f'(t) = \frac{1}{\operatorname{per} A_t} \sum_{i,j} (b_{ij} - a_{ij}) \operatorname{per} A_t(i|j),$$

while by Lemma 4.3 we have

$$\frac{\operatorname{per} A_t(i|j)}{\operatorname{per} A_t} \leqslant \frac{\lambda^4 n}{\sum_{i,j} (1-t) a_{ij} + t b_{ij}} \,.$$

Plugging these estimates into (4.6) we obtain the desired inequality. $\qquad \square$

As another consequence of Lemma 4.3, we obtain the following estimate on how the permanental mean of a matrix varies when a row and a column are deleted:

Lemma 4.5. *There is a constant $C > 1$ such that for any $\lambda > 1$ and any $n \geqslant 2$, if $A \in \mathbb{R}_{++}^{n \times n}$ is a λ-bounded matrix then for any i, $j \in \{1, \ldots, n\}$ we have*

$$\left| \log \frac{\operatorname{pm}(A(i|j))}{\operatorname{pm}(A)} \right| \leqslant \frac{6 \log \lambda + C \log n}{n} \,.$$

Proof. By Stirling's formula, $\log(n!) = n \log n - n + \mathcal{O}(\log n)$, and so

$$\log \operatorname{pm}(A) = \frac{1}{n} \log \operatorname{per} A + \log n - 1 + \mathcal{O}\left(\frac{\log n}{n}\right) .$$

Analogously, letting $B := A(i|j)$,

$$\log \operatorname{pm}(B) = \frac{1}{n-1} \log \operatorname{per} B + \log n - 1 + \mathcal{O}\left(\frac{\log n}{n}\right) .$$

These two estimates yield:

$$\left|\log \frac{\operatorname{pm}(B)}{\operatorname{pm}(A)}\right| \leqslant \left|\frac{1}{n} \log \operatorname{pm}(B)\right| + \left|\frac{n-1}{n} \log \operatorname{pm}(B) - \log \operatorname{pm}(A)\right|$$

$$\leqslant \underbrace{\left|\frac{1}{n} \log \operatorname{pm}(B)\right|}_{(\star)} + \underbrace{\left|\frac{1}{n} \log \frac{\operatorname{per} B}{\operatorname{per} A}\right|}_{(\star\star)} + \mathcal{O}\left(\frac{\log n}{n}\right) .$$

Since B is λ-bounded, it follows from internality of the permanental mean that $(\star) \leqslant (\log \lambda)/n$. On the other hand, by Lemma 4.3 we have $(\star\star) \leqslant (5 \log \lambda + \log n)/n$. The lemma follows. $\square$

4.4. **Proof of Theorem 4.1.**

4.4.0. *Zeroth step: Some reductions.* In order to prove Theorem 4.1, it is sufficient to consider functions that, in addition of being in $\mathcal{B}(\mu \times \nu)$, are doubly stochastic, i.e., satisfy relations (3.10). Indeed, assuming the theorem already proved in this case, and given an arbitrary $f \in \mathcal{B}(\mu \times \nu)$, we consider the Sinkhorn decomposition $f(x,y) = \varphi(x)g(x,y)\psi(y)$ as in (3.11) given by Theorem 3.6. Then, by row-wise and column-wise homogeneity of the permanental mean,

$$\operatorname{pm}\left(\square_n f(x,y)\right) = \left(\prod_{i=0}^{n-1} \varphi(T^i x)\right)^{1/n} \left(\prod_{j=0}^{n-1} \psi(S^j y)\right)^{1/n} \operatorname{pm}\left(\square_n g(x,y)\right) .$$

Since T is ergodic, by Birkhoff's theorem the first factor on the RHS converges to the geometric mean (3.1) for μ-a.e. $x \in X$. Analogously for the second factor. Since we are assuming Theorem 4.1 already proved in the doubly stochastic case, the third factor converges to the scaling mean of g, which by Proposition 3.3 equals 1. In conclusion, $\operatorname{pm}(\square_n f)$ converges a.e. to $\operatorname{gm}(\varphi)\operatorname{gm}(\psi)$, which by Proposition 3.5(a) equals $\operatorname{sm}(f)$.

So from this point on we assume that

$$f \in \mathcal{B}(\mu \times \nu) \text{ is doubly stochastic,} \tag{4.7}$$

and our aim is to show that the permanental means of the matrices (4.1) converge a.e. to 1.

As a second reduction, it is sufficient to prove Theorem 4.1 assuming that the measures μ and ν are non-atomic. Indeed, fix an arbitrary non-atomic Lebesgue probability space (Z, θ) and an ergodic measure preserving transformation $U \colon Z \to Z$, and consider the product spaces $(\hat{X}, \hat{\mu}) := (X \times Z, \mu \times \theta)$ and $(\hat{Y}, \hat{\nu}) := (Y \times Z, \nu \times \theta)$. Then the transformations $\hat{T}(x, z) := (Tx, Uz)$ and $\hat{S}(y, w) := (Sy, Uw)$ preserve the measures $\hat{\mu}$ and $\hat{\nu}$, respectively, and are ergodic. Given a doubly stochastic function $f \in (\log L^\infty)(\mu \times \nu)$, we consider the doubly stochastic function $\hat{f}(x, z, y, w) := f(x, y)$ on $\hat{X} \times \hat{Y}$. Since the measures $\hat{\mu}$ and $\hat{\nu}$ are non-atomic,

and assuming that Theorem 4.1 is already proved in this case, we conclude that the permanental mean of the matrix $\square_n \hat{f}(x, z, y, w)$ converges to 1 at $\hat{\mu} \times \hat{\nu}$-a.e. (x, z, y, w). But the latter matrix obviously equals $\square_n f(x, y)$, so we obtain (4.2).

So we can assume that the Lebesgue spaces (X, μ) and (Y, ν) are non-atomic. Actually, for convenience in the following proof, we will actually assume that

$$X = Y = [0, 1] \quad \text{and} \quad \mu = \nu \text{ is Lebesgue measure.} \tag{4.8}$$

4.4.1. *First step: Discretizing f.* Fix an arbitrary $\varepsilon > 0$. Let us construct a convenient discretized approximation of the given function f.

Let k be a large positive integer (to be specified later), and define a positive matrix $A = (a_{pq}) \in \mathbb{R}_{++}^{k \times k}$ by

$$a_{pq} := k^2 \int_{(p-1)/k}^{p/k} \int_{(q-1)/k}^{q/k} f(x, y) \, \mathrm{d}y \, \mathrm{d}x \,.$$

As a consequence of (4.7), the matrix $k^{-1}A$ is doubly stochastic. Indeed, for every $p \in \{1, \ldots, k\}$ the sum of the corresponding row of A is

$$\sum_{q=1}^{k} a_{pq} = k^2 \int_{(p-1)/k}^{p/k} \underbrace{\int_0^1 f(x, y) \, \mathrm{d}y}_{1} \, \mathrm{d}x = k \,,$$

and in an analogous way we compute the column sums. In particular, by the homogeneity of the scaling mean and Proposition 2.4, we have $\mathrm{sm}(A) = 1$.

Let g be a function on $X \times Y = [0, 1]^2$ equal to the constant a_{pq} on each sub-square $[(p-1)/k, p/k) \times [(q-1)/k, q/k)$, where $p, q \in \{1, \ldots, k\}$. Notice that g is nothing but the conditional expectation of f with respect to the σ-algebra generated by the partition into these sub-squares. Therefore, by Proposition 4.2, if k is chosen large enough then f and g are L^1-close in the sense that

$$\int |f - g| \, \mathrm{d}(\mu \times \nu) < \varepsilon. \tag{4.9}$$

We fix the integer k and therefore the matrix A and the function g from now on.

Since $f \in \mathcal{B}(\mu \times \nu)$, there exists $\lambda > 0$ such that

$$\lambda^{-1} \leqslant f \leqslant \lambda \qquad \mu \times \nu\text{-a.e.}$$

The values of g are obtained by averaging the values of f and therefore satisfy the same bounds, that is, $\lambda^{-1} \leqslant g \leqslant \lambda$. In particular, the matrices $\square_n f(x, y)$ and $\square_n g(x, y)$ are λ-bounded for a.e. (x, y). We now use Lemma 4.4 to compare their permanental means:

$$\left| \log \frac{\mathrm{pm}(\square_n f(x, y))}{\mathrm{pm}(\square_n g(x, y))} \right| = \frac{1}{n} \left| \log \frac{\mathrm{per}(\square_n f(x, y))}{\mathrm{per}(\square_n g(x, y))} \right| \leqslant \frac{\lambda^5}{n^2} \sum_{i,j=0}^{n} |f(T^i x, S^j y) - g(T^i x, S^j y)| \,.$$

Using the ergodic theorem (4.3) and the bound (4.9), we conclude that for a.e. (x, y),

$$\limsup_{n \to \infty} \left| \log \frac{\mathrm{pm}(\square_n f(x, y))}{\mathrm{pm}(\square_n g(x, y))} \right| < \lambda^5 \varepsilon \,. \tag{4.10}$$

4.4.2. *Second step: Comparing $\square_n g$ with a Kronecker product.* Let us fix (x, y) such that (4.10) holds, the T-orbit of x visits each of the intervals $[0, 1/k]$, $[1/k, 2/k]$, $\ldots$, $[(k-1)/k, 1]$ with limit frequency $1/k$, and analogously for the S-orbit of y.

For each $n \geqslant 1$, since the points x and y are not periodic, there exist permutation τ_n and σ_n of the set $\{0, 1, \ldots, n-1\}$ such that

$$T^{\tau_n(0)}x < T^{\tau_n(1)}x < \cdots < T^{\tau_n(n-1)}x, \qquad S^{\sigma_n(0)}y < S^{\sigma_n(1)}y < \cdots < S^{\sigma_n(n-1)}y.$$

Let P_{τ_n} be the $n \times n$ permutation matrix whose 1's are located on the positions $(i, \tau(i))$, $i \in \{0, 1, \ldots, n-1\}$. Analogously define P_{σ_n}. Consider the product matrix

$$B_n := P_{\tau_n} \cdot \square_n g(x, y) \cdot P_{\sigma_n}^{-1} = \left(g(T^{\tau_n(i)}x, S^{\sigma_n(j)}y)\right)_{i,j=0}^{n-1}.$$

This matrix has a decomposition into blocks:

$$B_n = \begin{pmatrix} B_{n,11} & B_{n,12} & \cdots & B_{n,1k} \\ B_{n,21} & B_{n,22} & \cdots & B_{n,2k} \\ \vdots & \vdots & & \vdots \\ B_{n,k1} & B_{n,k2} & \cdots & B_{n,kk} \end{pmatrix}$$

where the block $B_{n,pq}$ has all its entries are equal to a_{pq}, has width equal to the cardinality of $\{x, Tx, \cdots, T^{n-1}x\} \cap [(p-1)/k, p/k]$, and has height equal to the cardinality of $\{y, Sy, \cdots, S^{n-1}y\} \cap [(q-1)/k, q/k]$. It follows that if $n = km$ for some sufficiently large integer m then the matrix B_n and the Kronecker product $A \otimes U_m$ differ at at most εn rows and columns. Since both matrices are λ-bounded, by Lemma 4.4 we have

$$\left|\log \frac{\mathrm{pm}(B_n)}{\mathrm{pm}(A \otimes U_m)}\right| = \frac{1}{n}\left|\log \frac{\mathrm{per}(B_n)}{\mathrm{per}(A \otimes U_m)}\right| \leqslant \frac{1}{n}\frac{\lambda^5}{n}(\lambda - \lambda^{-1})\varepsilon n^2 < \lambda^6 \varepsilon.$$

Since the matrices B_n and $\square_n g(x, y)$ are by definition permutationally equivalent, they have the same permanental mean (by the row-/column-wise symmetry property). On the other hand, by the generalized Friedland limit (Theorem 2.19), we have

$$\lim_{m \to \infty} \mathrm{pm}(A \otimes U_m) = \mathrm{sm}(A) = 1.$$

So we obtain

$$\limsup_{m \to \infty} |\log \mathrm{pm}(\square_{km}g(x, y))| \leqslant \lambda^6 \varepsilon.$$

Notice that, as a consequence of Lemma 4.5,

$$\lim_{n \to \infty} \log \frac{\mathrm{pm}(\square_{k\lfloor n/k\rfloor}\, g(x, y))}{\mathrm{pm}(\square_n g(x, y))} = 0,$$

and so

$$\limsup_{n \to \infty} |\log \mathrm{pm}(\square_n g(x, y))| \leqslant \lambda^6 \varepsilon.$$

Recalling (4.10), we obtain

$$\limsup_{n \to \infty} |\log \mathrm{pm}(\square_n f(x, y))| \leqslant (\lambda^6 + \lambda^5)\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we infer that

$$\lim_{n \to \infty} \mathrm{pm}(\square_n f(x, y)) = 1.$$

This proves (4.2) under the assumptions (4.7) and (4.8). As explained in § 4.4.0, Theorem 4.1 in full generality follows.

5. Applications

5.1. **Symmetric means.** The *elementary symmetric polynomial* of degree k in $n \geqslant k$ variables is defined as

$$E_k(z_1, z_2, \ldots, z_n) := \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} z_{i_1} z_{i_2} \cdots z_{i_k}.$$

These sums appears in a wide range of different areas of mathematics. For example, Vieta's formula states that if $P(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ is a monic polynomial and $z_1, \ldots, z_n$ are its roots (repeated according to multiplicity) then $a_{n-k} = (-1)^k E_k(z_1, \ldots, z_n)$.

Assuming that $z = (z_1, \ldots, z_n)$ is a string of nonnegative real numbers, we define its *k-th symmetric mean* as

$$\mathrm{sym}_k(z) := \left(\frac{E_k(z_1, \ldots, z_n)}{\binom{n}{k}} \right)^{1/k}.$$

Notice that $\mathrm{sym}_1(z)$ is the arithmetic mean, and $\mathrm{sym}_n(z)$ is the geometric mean. In the 18th century, MacLaurin discovered that $\mathrm{sym}_1(z) \geqslant \mathrm{sym}_2(z) \geqslant \cdots \geqslant \mathrm{sym}_n(z)$, thus generalizing the AM–GM inequality (see [HLP, § 2.22]).

Symmetric means have the properties of reflexivity, monotonicity, internality, continuity, homogeneity, and, of course, symmetry. Actually, they can be expressed in terms of permanental means: we have

$$\mathrm{sym}_k(z) = \left[\mathrm{pm}(R_k(z)) \right]^{n/k}, \tag{5.1}$$

where

$$R_k(z_1, \ldots, z_n) := \left. \begin{pmatrix} z_1 & z_2 & \cdots & z_n \\ \vdots & \vdots & & \vdots \\ z_1 & z_2 & \cdots & z_n \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{array}{l} \left. \vphantom{\begin{matrix} z_1 \\ \vdots \\ z_1 \end{matrix}} \right\} k \text{ rows} \\ \left. \vphantom{\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}} \right\} n-k \text{ rows} \end{array} \tag{5.2}$$

Using this relation, we can deduce from the Law of Large Permanents (Theorem 4.1) an ergodic theorem for symmetric means. Actually, such result already exists and was obtained in 1976 by Halász and Székely [HS]:

Theorem 5.1 (Halász–Székely). *Let (X, μ) be a Lebesgue probability space, $T \colon X \to X$ be a measure-preserving ergodic transformation, $g \in \mathcal{B}(\mu)$, and $0 < c < 1$. Suppose $k(n)$ is a sequence of integers satisfying*

$$1 \leqslant k(n) \leqslant n \quad and \quad \lim_{n \to \infty} \frac{k(n)}{n} = c.$$

Then, for μ-almost every x, the limit

$$\lim_{n \to \infty} \mathrm{sym}_{k(n)} \left(g(x), g(Tx), \ldots, g(T^{n-1}x) \right)$$

exists and equals

$$\mathrm{sym}_c(g) := c \left(\frac{1-c}{r} \right)^{\frac{1-c}{c}} \exp \left(\frac{1}{c} \int \log(g + r) \, \mathrm{d}\mu \right), \tag{5.3}$$

where $r = r(c)$ is the unique positive root of the equation

$$\int \frac{g}{g + r} \, \mathrm{d}\mu = c \, .$$

Proof. Let $Y = \{0,1\}^{\mathbb{N}}$, let ν be the Bernoulli measure with weights c, $1 - c$, let $S \colon Y \to Y$ be the shift, and let $\{Y_0, Y_1\}$ the partition of Y into the cylinders of length 1. Consider the function

$$f(x, y) := \begin{cases} g(x) & \text{if } y \in Y_0 \, , \\ 1 & \text{if } y \in Y_1 \, . \end{cases}$$

For μ-a.e. x and ν-a.e. y, the conclusion (4.2) of Theorem 4.1 holds, and moreover if $\ell(n)$ denotes the cardinality of the set $\{y, Sy, \ldots, S^{n-1}y\} \cap Y_0$ then $\ell(n)/n \to c$ as $n \to \infty$. Using notation (5.2), define matrices

$$A_n := R_{\ell(n)}(g(x), \ldots, g(T^{n-1}x)) \, , \qquad B_n := R_{k(n)}(g(x), \ldots, g(T^{n-1}x)) \, .$$

Notice that A_n can be obtained from $\square_n f(x, y)$ by permuting rows. In particular,

$$\lim_{n \to \infty} \mathrm{pm}(A_n) = \lim_{n \to \infty} \mathrm{pm}(\square_n f(x, y)) = \mathrm{sm}(f) \, .$$

Let $\lambda > 1$ be such that $\lambda^{-1} \leqslant g \leqslant \lambda$. Then the matrices A_n and B_n are λ-bounded, and so by Lemma 4.4,

$$\left| \log \frac{\mathrm{pm}(B_n)}{\mathrm{pm}(A_n)} \right| = \frac{1}{n} \left| \log \frac{\mathrm{per}(B_n)}{\mathrm{per}(A_n)} \right| \leqslant \frac{1}{n} \frac{\lambda^5}{n} \left(\lambda - \lambda^{-1} \right) n |\ell(n) - k(n)| \leqslant \frac{\lambda^6 |\ell(n) - k(n)|}{n} \, ,$$

which converges to 0 as $n \to \infty$. So $\mathrm{pm}(B_n) \to \mathrm{sm}(f)$, and therefore by (5.1) we conclude that $\mathrm{sym}_{k(n)}(g(x), \ldots, g(T^{n-1}x)) \to [\mathrm{sm}(f)]^{1/c}$. We conclude the proof using Proposition 3.9 to compute $\mathrm{sm}(f)$. $\qquad \square$

Let us compare the result above with that of Halász and Székely's paper [HS]. The theorem from that paper requires only a weak integrability condition, namely, $\int \log(1 + g) \, \mathrm{d}\mu < \infty$. The theorem is stated in terms of independent identically distributed random variables, but the proof actually does not use independence, and ergodicity suffices. Therefore, the actual Halász–Székely theorem is stronger than Theorem 5.1 above. This indicates that a weakening of the hypotheses of Theorem 4.1 should be pursued (more about this on Section 6 below) and should not be regarded as an inherent drawback of our approach. Let us mention that the proof in [HS] employs completely different tools, namely: Vieta's formula and Cauchy integral formula are used to relate the means with a certain complex integral, and then the saddle point method is used to estimate the value of the integral. This line of argument has been used in most of the probability papers on the subject. Our methods, on the other hand, provide a more transparent explanation for the complicated formula (5.3), and apply to much more general types of means. A second example of application is given in the next subsection.

5.2. Muirhead means. Let $z = (z_1, \ldots, z_n)$ and $\alpha = (\alpha_1, \ldots, \alpha_n)$ be two strings of nonnegative numbers, not all the α_i's being 0. We then define the *Muirhead α-mean* of the z_i's as

$$\mathrm{mum}_\alpha(z) := \left(\frac{1}{n!} \sum_{\sigma \in S_n} \prod_{i=1}^m z_{\sigma(i)}^{\alpha_i} \right)^{\frac{1}{\alpha_1 + \cdots + \alpha_n}} \, ,$$

where S_n denotes the set of permutations of $\{1, \ldots, n\}$. Note that this coincides with the arithmetic mean if $\alpha = (1, 0, \ldots, 0)$, and with the geometric mean if $\alpha = (1, 1, \ldots, 1)$. Muirhead means also generalize the symmetric ones; indeed

$$\mathrm{sym}_k(z) = \mathrm{mum}_{\alpha_k}(z) \quad \text{where} \quad \alpha_k := \big(\underbrace{1, 1, \ldots, 1}_{k}, \underbrace{0, 0, \ldots, 0}_{n-k}\big).$$

The celebrated Muirhead–Hardy–Littlewood–Pólya inequality [HLP, § 2.18] describes when the functions $\mathrm{mum}_\alpha(\cdot)$ and $\mathrm{mum}_\beta(\cdot)$ are comparable; this is done in terms of a concept called *majorization*, which appears in a vast number of other situations.

Muirhead means can also be expressed in terms of permanental ones:

$$\mathrm{mum}_\alpha(z) = [\mathrm{pm}(M_\alpha(z))]^{\frac{n}{\alpha_1 + \cdots + \alpha_n}} \quad \text{where} \quad M_\alpha(z) := \begin{pmatrix} z_1^{\alpha_1} & \cdots & z_n^{\alpha_1} \\ \vdots & & \vdots \\ z_1^{\alpha_n} & \cdots & z_n^{\alpha_n} \end{pmatrix}.$$

Therefore we can deduce from the Law of Large Permanents an ergodic theorem for Muirhead means, namely:

Corollary 5.2. *Let* (X, μ), (Y, ν) *be a Lebesgue probability spaces,* $T \colon X \to X$, $S \colon Y \to Y$ *be measure-preserving ergodic transformations, and* $g \in \mathcal{B}(\mu)$, $h \in \mathcal{B}(\nu)$. *Then, for* $\mu \times \nu$*-almost every* (x, y), *the limit*

$$\lim_{n \to \infty} \mathrm{mum}_{(h(y), \ldots, h(S^{n-1}y))} \big(g(x), g(Tx), \ldots, g(T^{n-1}x)\big)$$

exists and equals

$$\mathrm{mum}_h(g) := \big[\mathrm{sm}\big(g^h\big)\big]^{1/\int h \, d\nu}. \tag{5.4}$$

Remark 5.3. Formulas (5.3) and (5.4) can be though as "continuous" (i.e., functional) versions of the symmetric and the Muirhead means, respectively.

6. Open questions and directions for future research

Permanents and Sinkhorn decompositions also make sense for *multidimensional* matrices and functions (see e.g. [Ba1]). We believe that most of the results of this paper can be extended accordingly, but we have not checked the details.

After having obtained a law of large numbers, a natural step is to look for a central limit theorem. In the particular case of symmetric means covered by [HS] and Theorem 5.1 above, a central limit theorem was obtained by Székely [Sz], under the assumption of independence.

Concerning the Law of Large Permanents itself, we do not believe that the statement of Theorem 4.1 is the optimal one. We would like to weaken the hypothesis that f is (essentially) bounded away from 0 and ∞. This assumption was used twice in the proof: first, to apply Theorem 3.6, and second, to use the regularity estimates for the permanent from § 4.3. So in order to strengthen the Law of Large Permanents we will probably need to solve other problems which are themselves of independent interest:

- To obtain explicit necessary and sufficient conditions on the function f for the existence of a functional Sinkhorn decomposition $f = \varphi g \psi$ in the sense of Definition 3.4, or at least in the more restricted setting of direct products (§ 3.5).

- To improve the estimates from § 4.3 so that in particular we are able to deal with matrices containing zero entries.

The last and maybe most interesting line of research motivated by the results of this paper is to extend the Law of Large Permanents to infinite matrices whose entries form a $\mathbb{N}^2$-indexed stochastic process (or, in more dynamical terms, whose entries are the values of a given function along the orbit of an ergodic $\mathbb{N}^2$-action). Let us be more precise.

Suppose that T is an ergodic measure-preserving action of the semigroup $\mathbb{N}^2$ on a Lebesgue probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Given a function $f : \Omega \to \mathbb{R}$ we define the $n \times n$ matrix $\square_n f(\omega) := f(T^{(i,j)}(\omega))_{0 \leqslant i,j \leqslant n-1}$. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be the sub-σ-algebras formed by the $T^{(1,0)}$-invariant and the $T^{(0,1)}$-invariant sets, respectively. We believe that the following statement should hold:

Conjecture 6.1. If $\log f \in L^\infty(\mathbb{P})$ then for $\mathbb{P}$-a.e. ω,

$$\lim_{n \to \infty} \mathrm{pm} \left(\square_n f(\omega) \right) = \mathrm{sm}_{\mathcal{A}_1, \mathcal{A}_2}(f).$$

Conjecture 6.1 would follow from the next statement regarding sequences of matrices:

Conjecture 6.2. Let (A_n) be a sequence of matrices of increasing sizes $n \times n$, all with row and column arithmetic means equal to 1 (i.e. $\frac{1}{n} A_n$ is doubly stochastic for each n.) Suppose that there exists $\lambda > 1$ each matrix A_n is λ-bounded in the sense (4.5). Then

$$\lim_{n \to \infty} \mathrm{pm}(A_n) = 1.$$

It turns out that there are at least two claims in the literature at least as strong as Conjecture 6.2: see the references [G] and [Mc]. Unfortunately we are unable to verify the correctness of either.

References

[Ba1] BAPAT, R.B. $D_1 A D_2$ theorems for multidimensional matrices. *Linear Algebra Appl.* 48 (1982), 437442. (Cited on page 26.)

[Ba2] BAPAT, R.B. Applications of an inequality in information theory to matrices. *Linear Algebra Appl.* 78 (1986), 107–117. (Cited on page 9.)

[Ba3] BAPAT, R.B. Permanents in probability and statistics. *Linear Algebra Appl.* 127 (1990), 3–25. (Cited on page 3.)

[Bh] BHATIA, R. *Matrix analysis.* Graduate Texts in Mathematics, 169. Springer-Verlag, New York, 1997. (Cited on pages 6 and 12.)

[Bi] BIRKHOFF, G. Extensions of Jentzsch's theorem. *Trans. Amer. Math. Soc.* 85 (1957), 219–227. (Cited on page 16.)

[Bo] BOGACHEV, V.I. *Measure theory.* Vol. 2. Springer-Verlag, Berlin, 2007. (Cited on pages 13 and 19.)

[BLN] BORWEIN, J.M.; LEWIS, A.S.; NUSSBAUM, R.D. Entropy minimization, DAD problems, and doubly stochastic kernels. *J. Funct. Anal.* 123 (1994), no. 2, 264–307. (Cited on pages 15 and 17.)

[Br] BRUALDI, R.A. Permanent of the direct product of matrices. *Pacific J. Math.* 16 (1966), 471–482. (Cited on page 10.)

[BPS]   BRUALDI, R.A.; PARTER, S.V.; SCHNEIDER, H. The diagonal equivalence of a nonnegative matrix to a stochastic matrix. *J. Math. Anal. Appl.* 16 (1966), 31–50. (Cited on page 8.)

[BR]   BRUALDI, R.A.; RYSER, H.J. *Combinatorial matrix theory.* Encyclopedia of Mathematics and its Applications, 39. Cambridge University Press, Cambridge, 1991. (Cited on page 4.)

[Dj]   DJOKOVIĆ, D.Ž. Note on nonnegative matrices. *Proc. Amer. Math. Soc.* 25 (1970), 80–82. (Cited on page 9.)

[Eg]   EGORYCHEV, G.P. The solution of van der Waerden's problem for permanents. *Adv. in Math.* 42 (1981), no. 3, 299–305. (Cited on page 4.)

[Fa]   FALIKMAN, D.I. Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix. *Mat. Zametki* 29 (1981), no. 6, 931–938, 957. (Cited on page 4.)

[FL]   FRANKLIN, J.; LORENZ, J. On the scaling of multidimensional matrices. *Linear Algebra Appl.* 114/115 (1989), 717–735. (Cited on page 15.)

[Fr]   FRIEDLAND, S. A lower bound for the permanent of a doubly stochastic matrix. *Ann. of Math.* 110 (1979), no. 1, 167–176. (Cited on page 11.)

[FLS]   FRIEDLAND, S.; LI, C.-K.; SCHNEIDER, H. Additive decomposition of nonnegative matrices with applications to permanents and scaling. *Linear and Multilinear Algebra* 23 (1988), no. 1, 63–78. (Cited on page 12.)

[G]   GIRKO, V. L. The method of random determinants for estimating the permanent. *Random Oper. Stochastic Equations* 3 (1995), no. 2, 181–192. (Cited on page 27.)

[GvL]   GOLUB, G.H.; VAN LOAN, C.F. *Matrix computations.* Fourth edition. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2013. (Cited on page 7.)

[GS]   GURVITS, L.; SAMORODNITSKY, A. Bounds on the permanent and some applications. Preprint arXiv:1408.0976 (Cited on pages 10 and 12.)

[HS]   HALÁSZ, G; SZÉKELY, G.J. On the elementary symmetric polynomials of independent random variables. *Acta Math. Acad. Sci. Hungar.* 28 (1976), no. 3–4, 397–400. (Cited on pages 3, 24, 25, and 26.)

[HLP]   HARDY, G.H.; LITTLEWOOD, J.E.; PÓLYA, G. *Inequalities.* Reprint of the 1952 edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. (Cited on pages 24 and 26.)

[Ke]   KELLER, G. *Equilibrium states in ergodic theory.* London Mathematical Society Student Texts, 42. Cambridge University Press, Cambridge, 1998. (Cited on page 19.)

[KS]   KNOPP, P.; SINKHORN, R. A note concerning simultaneous integral equations. *Canad. J. Math.* 20 (1968), 855–861. (Cited on page 17.)

[Kr]   KRENGEL, U. *Ergodic theorems.* With a supplement by Antoine Brunel. de Gruyter Studies in Mathematics, 6. Walter de Gruyter and Co., Berlin, 1985. (Cited on page 19.)

[LS]   LAMOND, B.; STEWART, N.F. Bregman's balancing method. *Transportation Res. Part B* 15 (1981), no. 4, 239–248. (Cited on page 7.)

[LSW]   LINIAL, N.; SAMORODNITSKY, A.; WIGDERSON, A. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica* 20 (2000), no. 4, 545–568. (Cited on page 10.)

[Li]   LIVERANI, C. Decay of correlations. *Ann. of Math.* 142 (1995), no. 2, 239–301. (Cited on page 16.)

[Lo]   LONDON, D. On matrices with a doubly stochastic pattern. *J. Math. Anal. Appl.* 34 (1971), 648–652. (Cited on pages 7 and 9.)

[MM]   MARCUS, M.; MINC, H. *A survey of matrix theory and matrix inequalities.* Reprint of the 1969 edition. Dover Publications, Inc., New York, 1992. (Cited on pages 4, 9, and 10.)

[MO]   MARSHALL, A.W.; OLKIN, I. Scaling of matrices to achieve specified row and column sums. *Numer. Math.* 12 (1968), 83–90. (Cited on page 8.)

[Mc]   MCCULLAGH, P. An asymptotic approximation for the permanent of a doubly stochastic matrix. *J. Stat. Comput. Simul.* 84 (2014), no. 2, 404–414. (Cited on page 27.)

[Me1]   MENON, M.V. Reduction of a matrix with positive elements to a doubly stochastic matrix. *Proc. Amer. Math. Soc.* 18 (1967), 244–247. (Cited on page 15.)

[Me2]   MENON, M.V. Matrix links, an extremization problem, and the reduction of a non-negative matrix to one with prescribed row and column sums. *Canad. J. Math.* 20 (1968), 225–232. (Cited on page 8.)

[Mi]   MINC, H. *Permanents.* Encyclopedia of Mathematics and its Applications, Vol. 6. Encyclopedia of Mathematics and its Applications, 9999. Addison-Wesley Publishing Co., Reading, Mass., 1978. (Cited on pages 3, 4, and 9.)

[Nu1]   NUSSBAUM, R.D. Iterated nonlinear maps and Hilbert's projective metric. II. *Mem. Amer. Math. Soc.* 79 (1989), no. 401. iv+118 pp. (Cited on pages 15 and 16.)

[Nu2]   NUSSBAUM, R.D. Entropy minimization, Hilbert's projective metric, and scaling integral kernels. *J. Funct. Anal.* 115 (1993), no. 1, 45–99. (Cited on pages 15 and 17.)

[PM]   PERFECT, H.; MIRSKY, L. The distribution of positive elements in doubly-stochastic matrices. *J. Lond. Math. Soc.* 40 (1965), 689–698. (Cited on page 8.)

[RW]   REMPAŁA, G.; WESOŁOWSKI, J.. *Symmetric functionals on random matrices and random matchings problems.* Springer, New York, NY, 2008. (Cited on page 3.)

[Ro]   ROCKAFELLAR, R.T. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970 (Cited on page 5.)

[Si1]   SINKHORN, R. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.* 35 (1964), 876–879. (Cited on page 7.)

[Si2]   SINKHORN, R. Continuous dependence on A in the $D_1 A D_2$ theorems. *Proc. Amer. Math. Soc.* 32 (1972), 395–398. (Cited on page 8.)

[SK]   SINKHORN, R.; KNOPP, P. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.* 21 (1967), 343–348. (Cited on page 8.)

[Sz]   SZÉKELY, G.J. A limit theorem for elementary symmetric polynomials of independent random variables. *Z. Wahrsch. Verw. Gebiete* 59 (1982), no. 3, 355–359. (Cited on page 26.)

[TV]   TAO, T.; VU, V. On the permanent of random Bernoulli matrices. *Adv. Math.* 220 (2009), no. 3, 657–669. (Cited on page 3.)

[vdW]   VAN DER WAERDEN, B.L. Aufgabe 45. *Jber. Deutsch. Math. Verein.* 35 (1926), 117. (Cited on page 4.)

[vLW]   VAN LINT, J.H.; WILSON, R.M. *A course in combinatorics.* Cambridge University Press, Cambridge, 1992. (Cited on page 3.)

JAIRO BOCHI   GODOFREDO IOMMI   MARIO PONCE
jairo.bochi@mat.puc.cl   giommi@mat.puc.cl   mponcea@mat.puc.cl
www.mat.uc.cl/~jairo.bochi   www.mat.uc.cl/~giommi   www.mat.uc.cl/~mponcea

FACULTAD DE MATEMÁTICAS, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
AVENIDA VICUÑA MACKENNA 4860, SANTIAGO, CHILE